Title: Machine Learning (2021/2022)

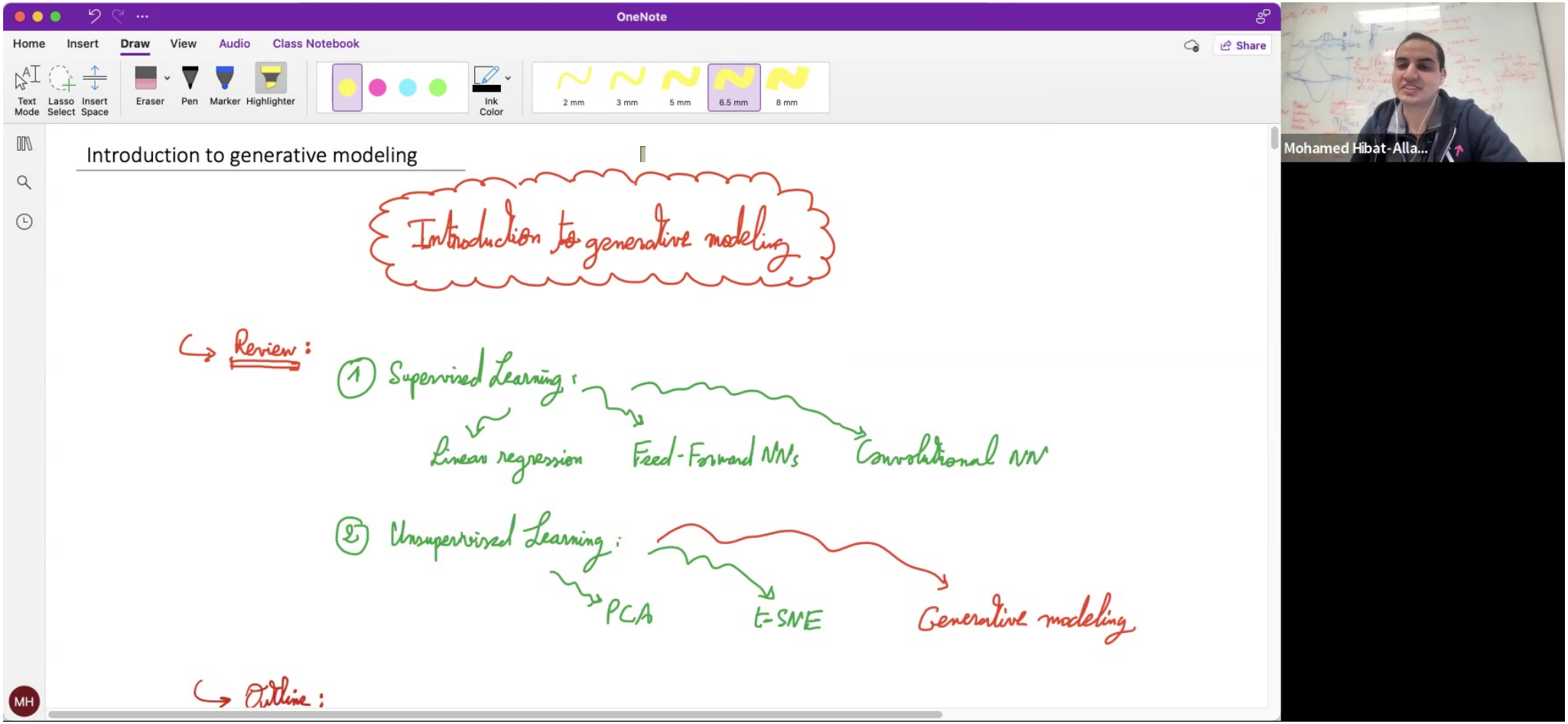
Speakers: Lauren Hayward

Collection: Machine Learning (2021/2022)

Date: April 26, 2022 - 11:30 AM

URL: https://pirsa.org/22040074

Abstract: This course is designed to introduce modern machine learning techniques for studying classical and quantum many-body problems encountered in condensed matter, quantum information, and related fields of physics. Lectures will focus on introducing machine learning algorithms and discussing how they can be applied to solve problem in statistical physics. Tutorials and homework assignments will concentrate on developing programming skills to study the problems presented in lecture.


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
Introduction to generative modeling
Introduction to generative modeling
Review:
1 Supervised Learning:
Linear regression
Feed-Forward NNs
Convolutional NN
2 Unsupervised Learning:
PCA
t-SNE
Generative modeling
Outline:
MH
Mohamed Hibat-Alla...

OneNote

Home Insert Draw View Audio Class Notebook

Share

Text Mode | Lasso Select | Insert Space | Eraser | Pen | Marker | Highlighter | Ink Color | 2 mm | 3 mm | 5 mm | 6.5 mm | 8 mm

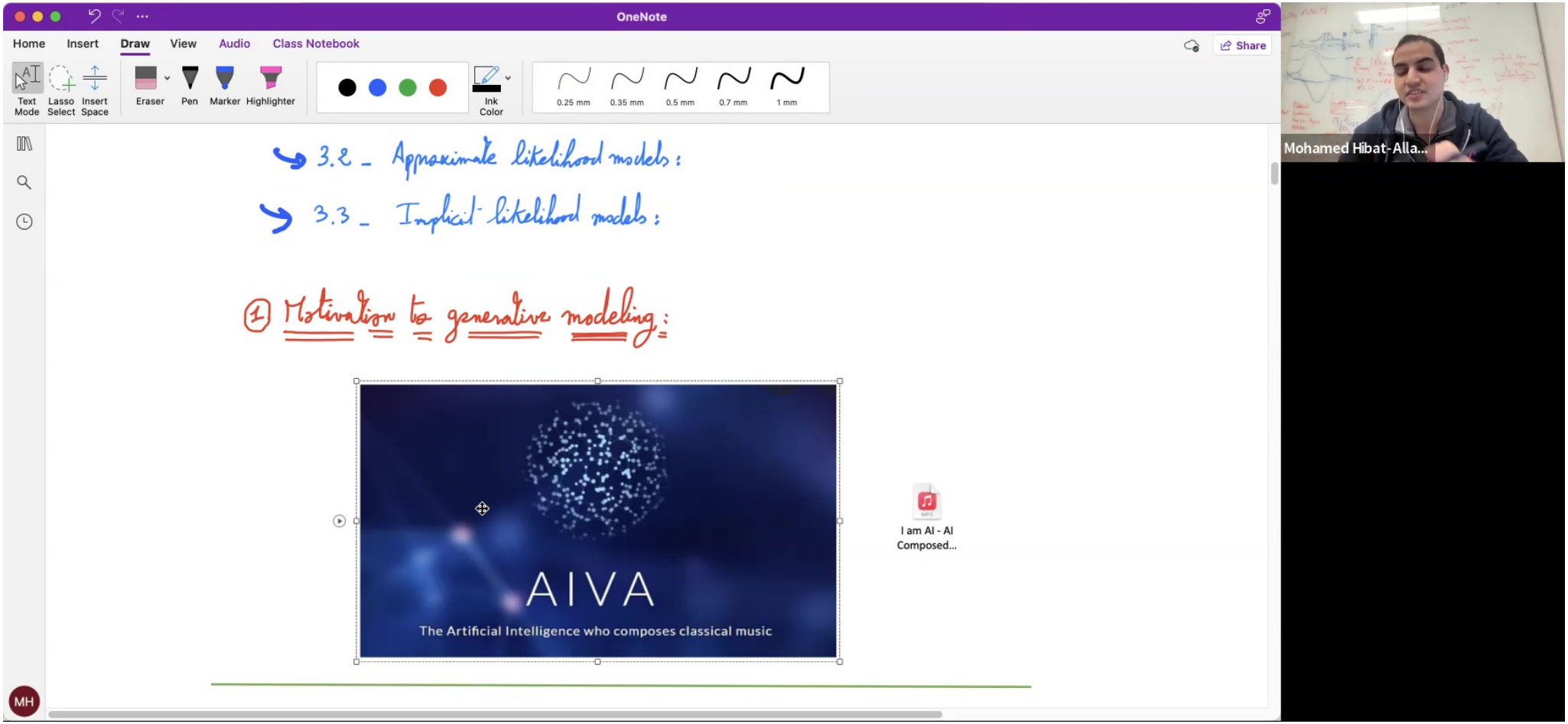
Mohamed Hibat-Alla...

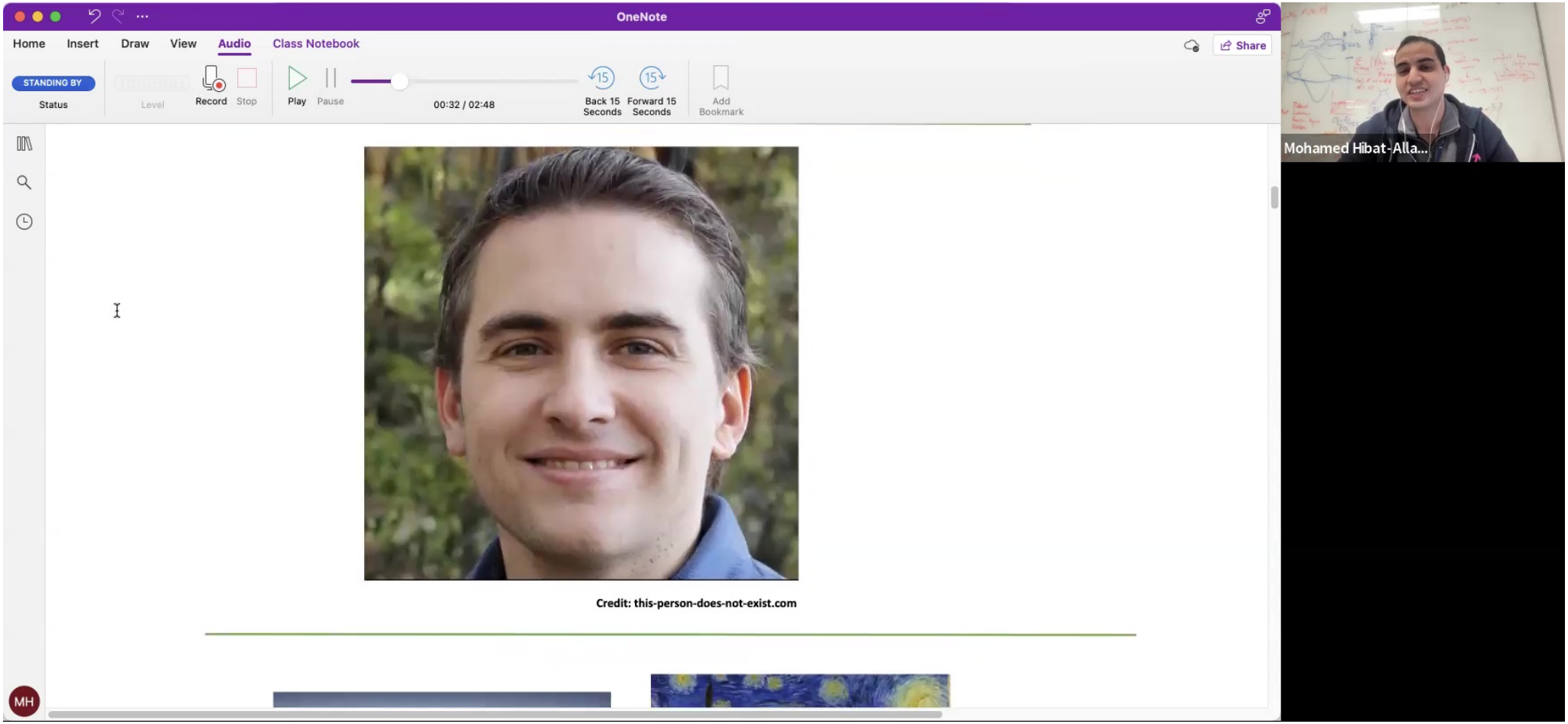
Outline:

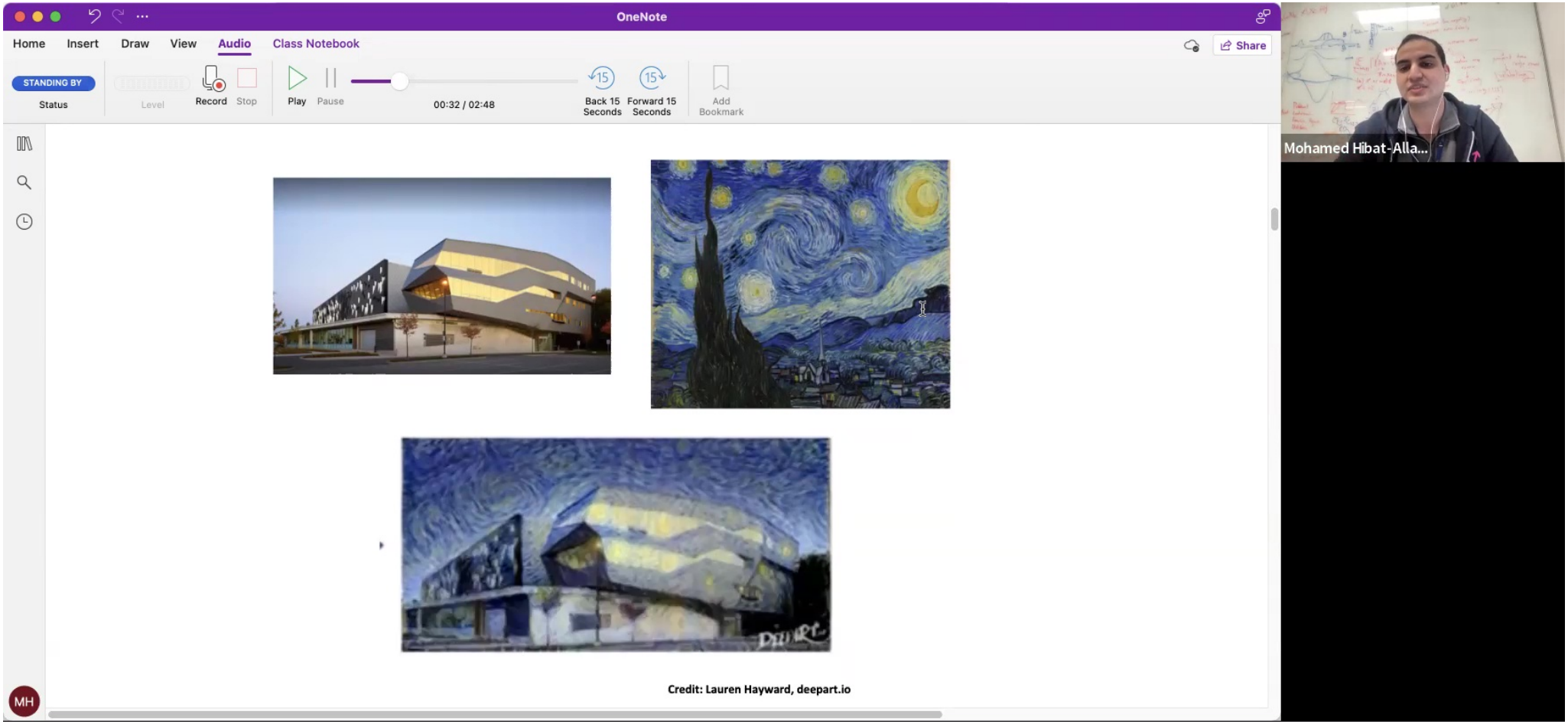
1. Motivation for generative modeling:
2. Maximum likelihood Principle & KL divergence:
3. Classification of generative models
   - 3.1 - Exact likelihood models:
   - 3.2 - Approximate likelihood models:
   - 3.3 - Implicit likelihood models:

1) Motivation to generative modeling:

MH


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
3.2 _ Approximate likelihood models:
3.3 _ Implicit likelihood models:
1 Motivation to generative modeling:
AIVA
The Artificial Intelligence who composes classical music
I am AI - AI Composed...
Mohamed Hibat-Alla...
MH


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
STANDING BY
Status
Level
Record
Stop
Play
Pause
00:32 / 02:48
Back 15 Seconds
Forward 15 Seconds
Add Bookmark
Credit: this-person-does-not-exist.com
MH
Mohamed Hibat-Alla...


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
STANDING BY
Status
Level
Record
Stop
Play
Pause
00:32 / 02:48
Back 15 Seconds
Forward 15 Seconds
Add Bookmark
Mohamed Hibat-Alla...
Credit: Lauren Hayward, deepart.io
MH

OneNote

Home Insert Draw View Audio Class Notebook

Share

STANDING BY
Status

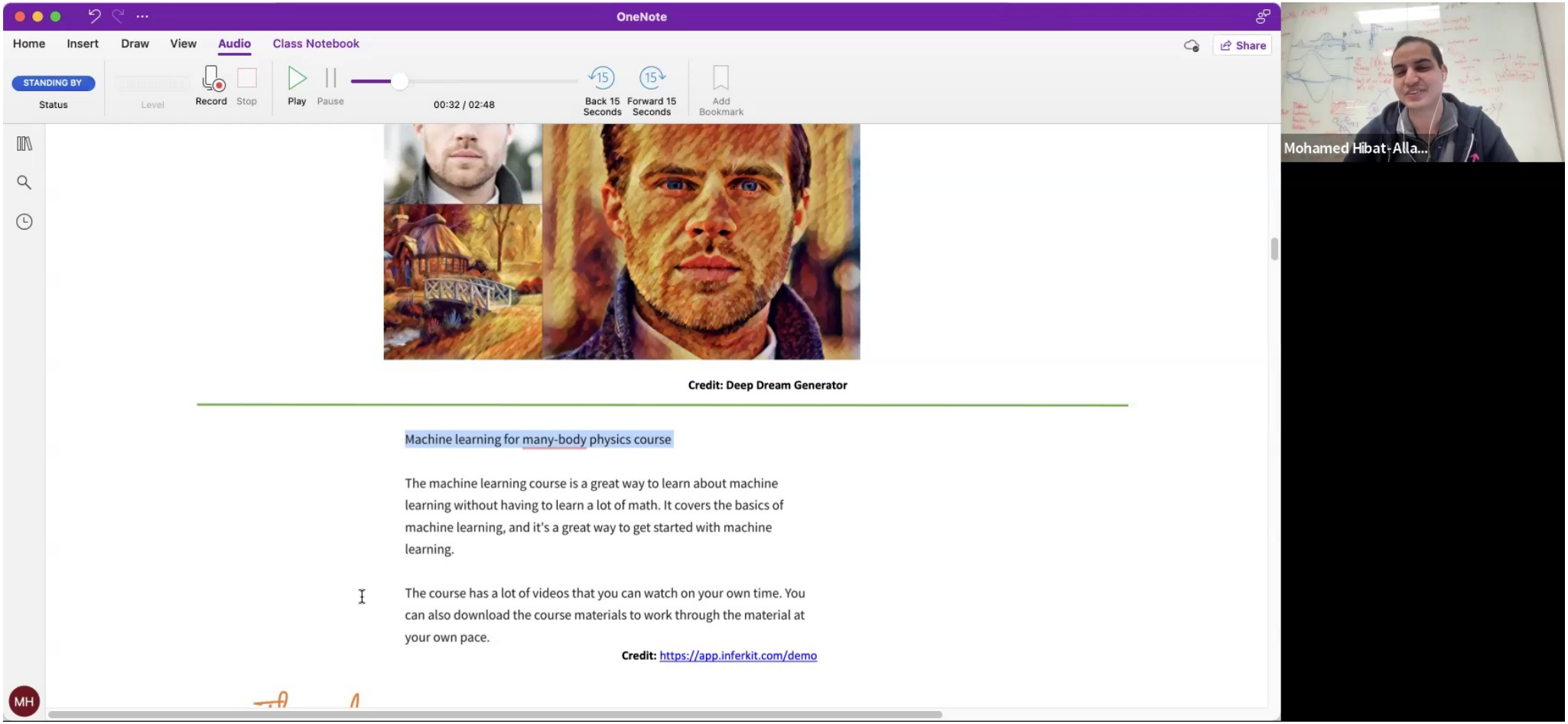
Level Record Stop Play Pause 00:32 / 02:48 Back 15 Seconds Forward 15 Seconds Add Bookmark

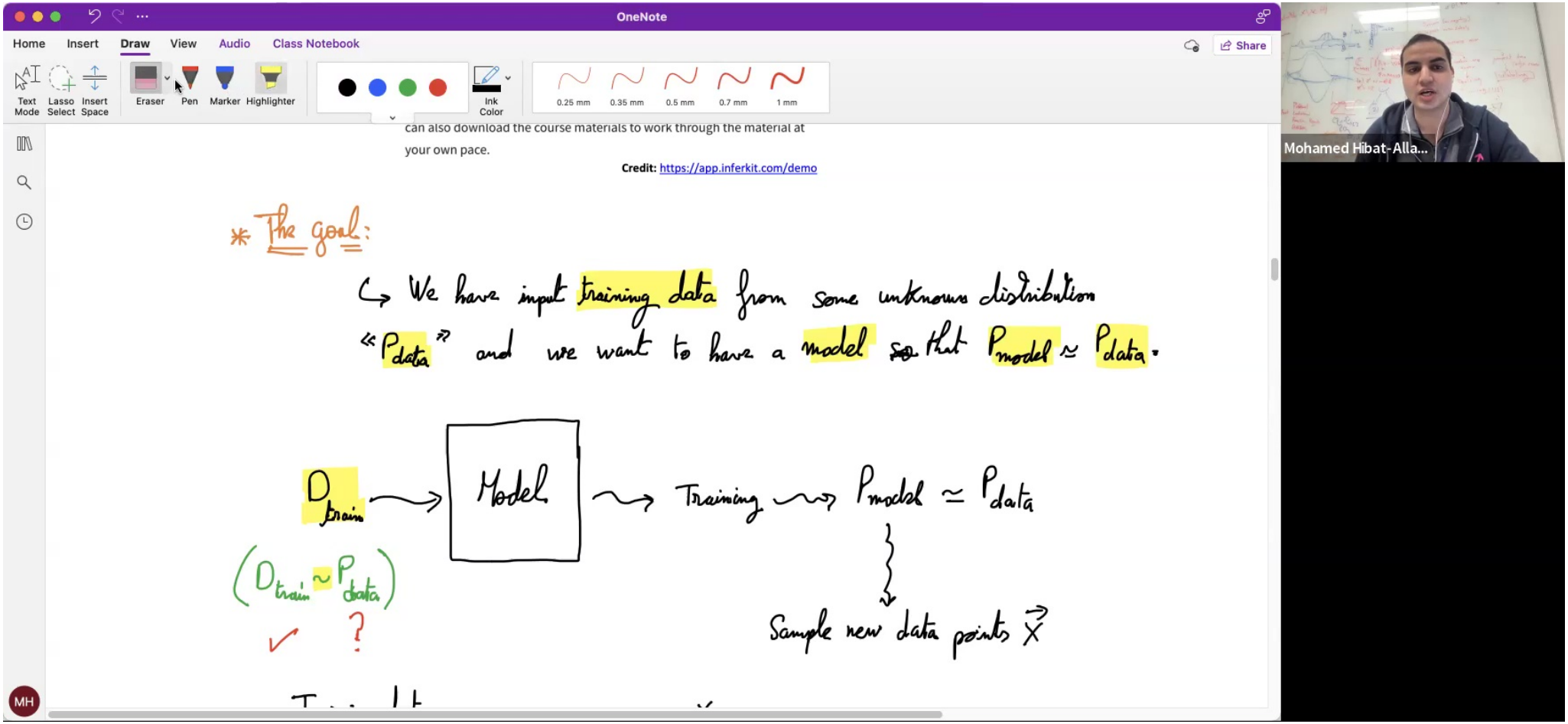
Credit: Lauren Hayward, deepart.io

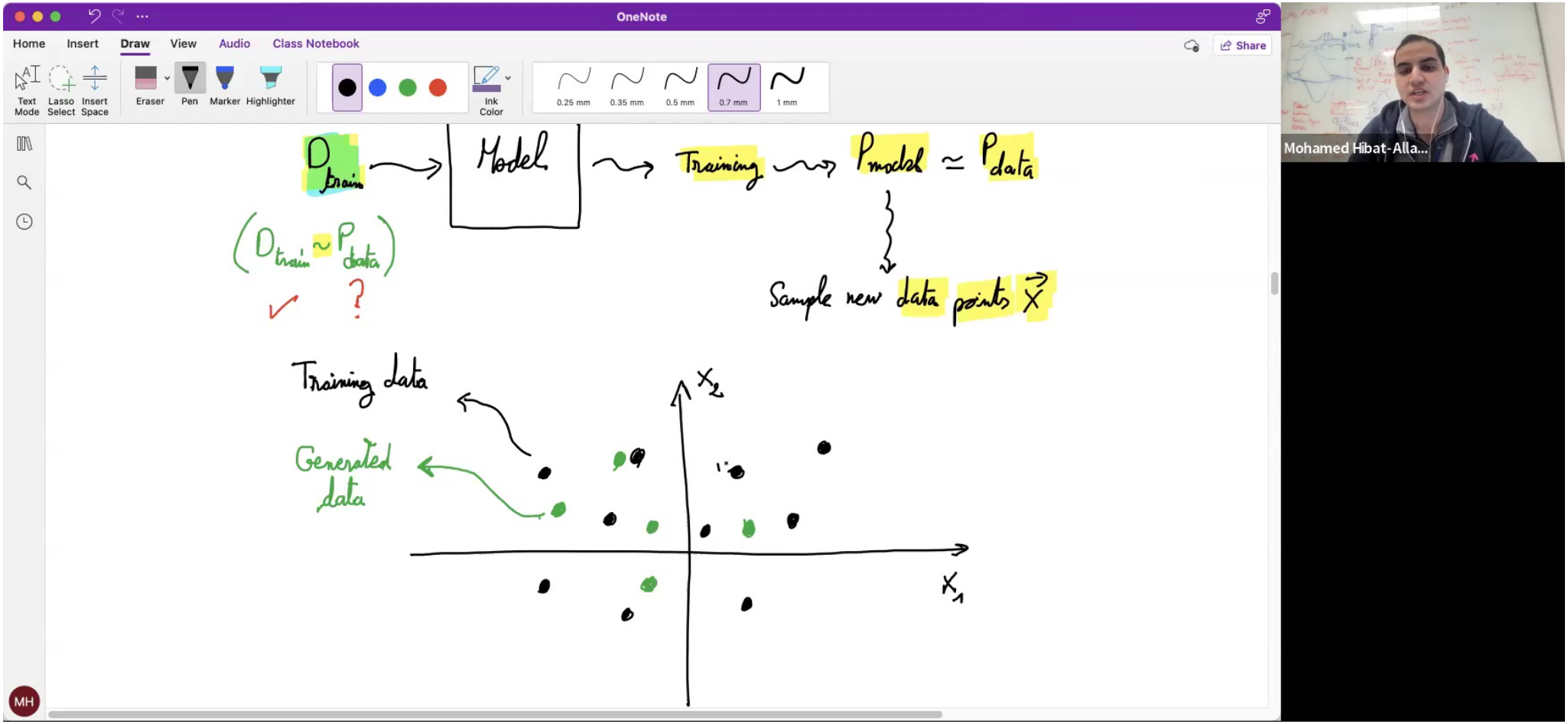
Credit: Deep Dream Generator

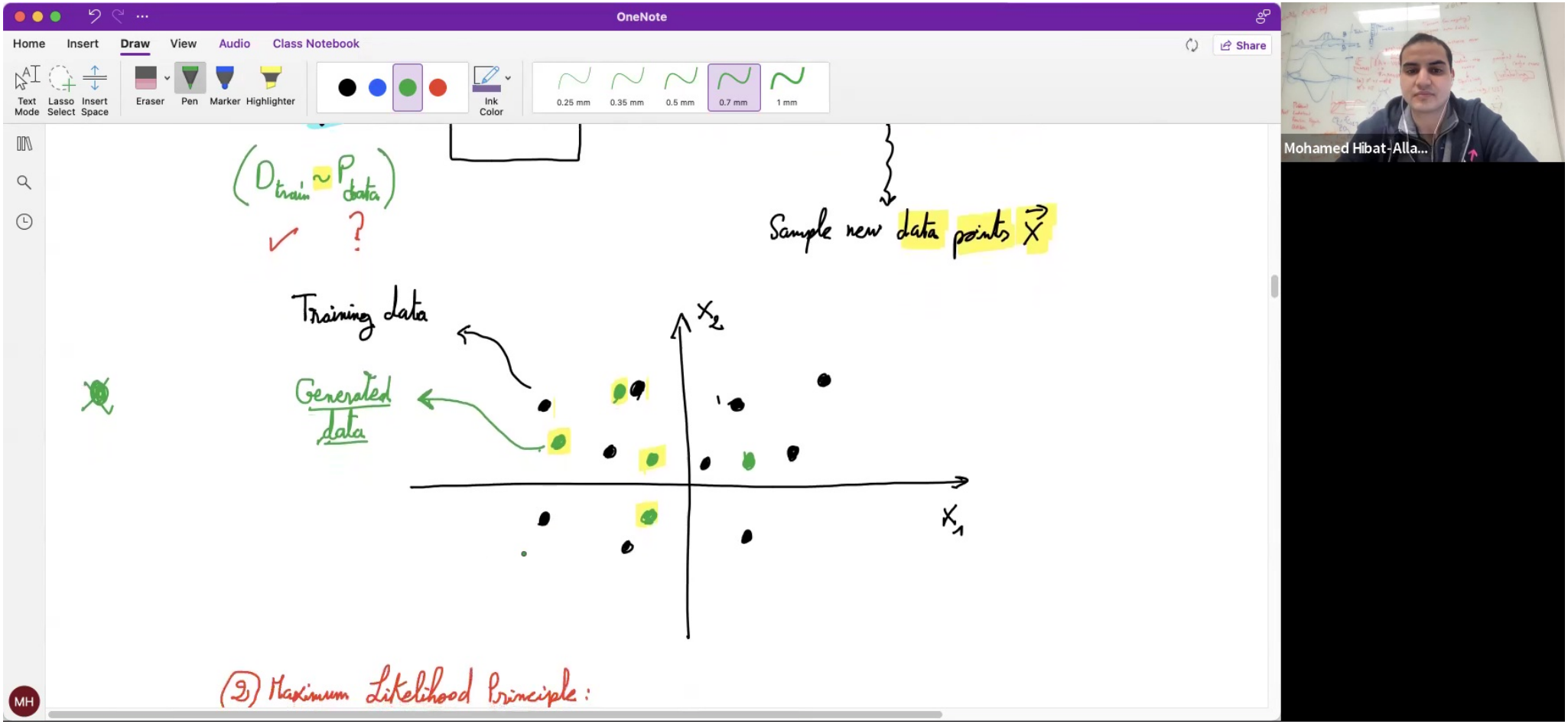
MH

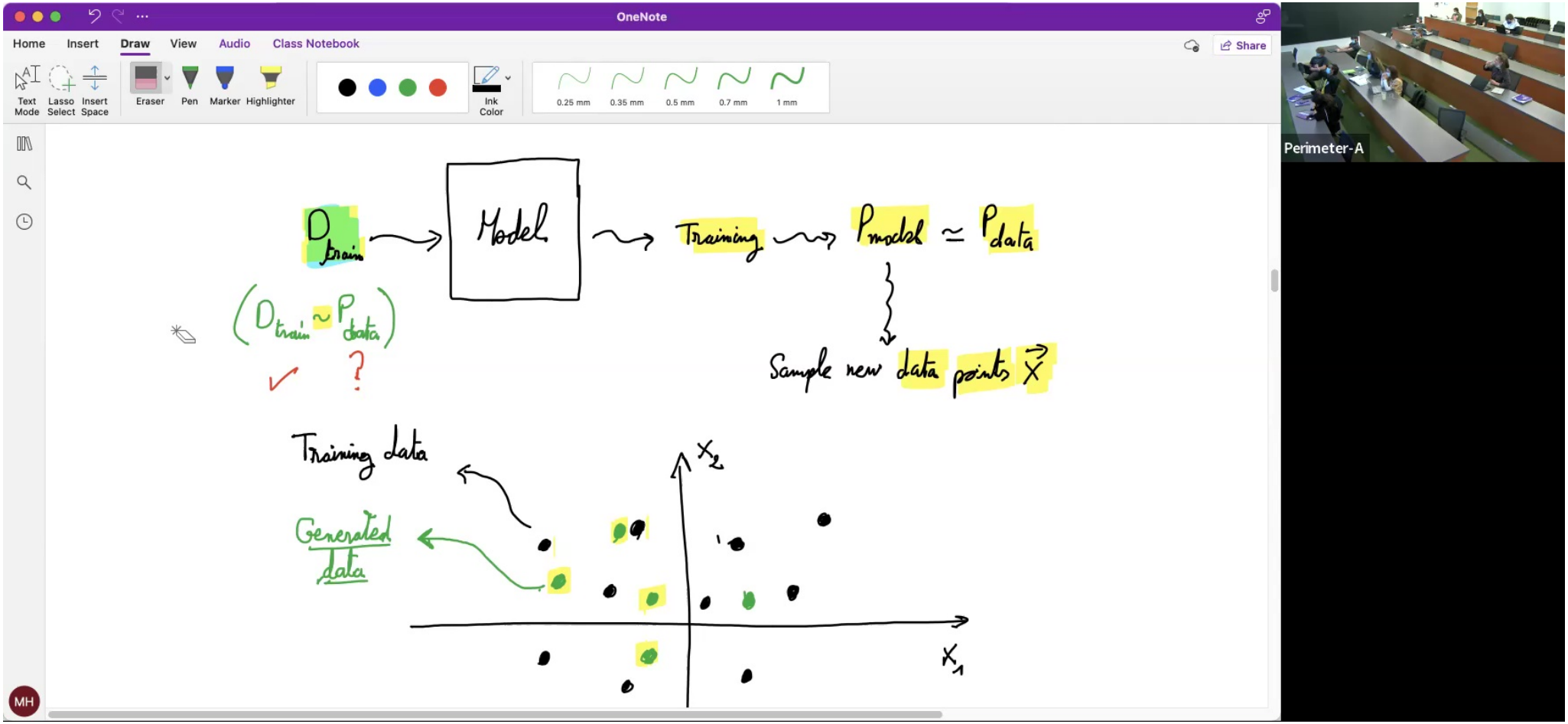
Mohamed Hibat-Alla...

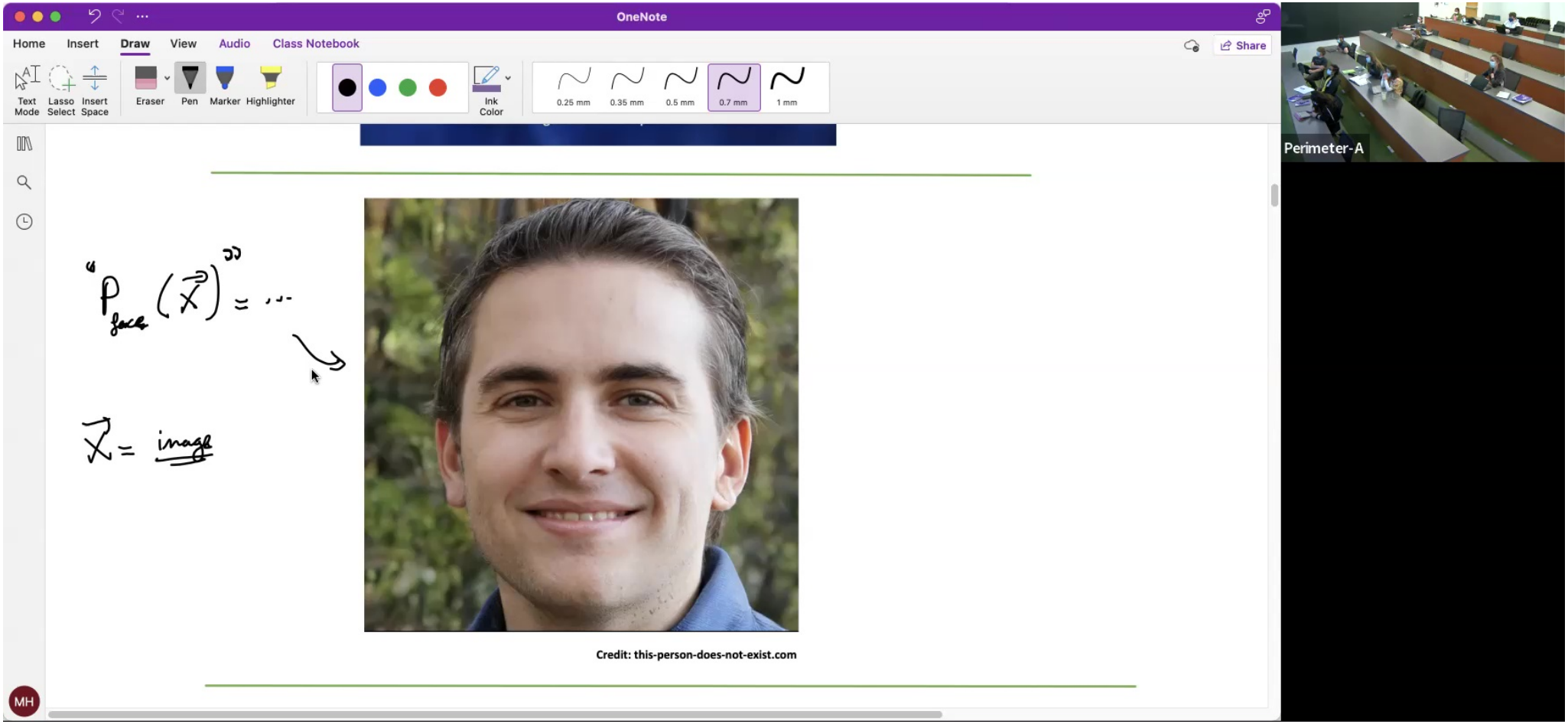

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
STANDING BY
Status
Level
Record
Stop
Play
Pause
00:32 / 02:48
Back 15 Seconds
Forward 15 Seconds
Add Bookmark
Credit: Deep Dream Generator
Machine learning for many-body physics course
The machine learning course is a great way to learn about machine learning without having to learn a lot of math. It covers the basics of machine learning, and it's a great way to get started with machine learning.
The course has a lot of videos that you can watch on your own time. You can also download the course materials to work through the material at your own pace.
Credit: https://app.inferkit.com/demo
MH
Mohamed Hibat-Alla...

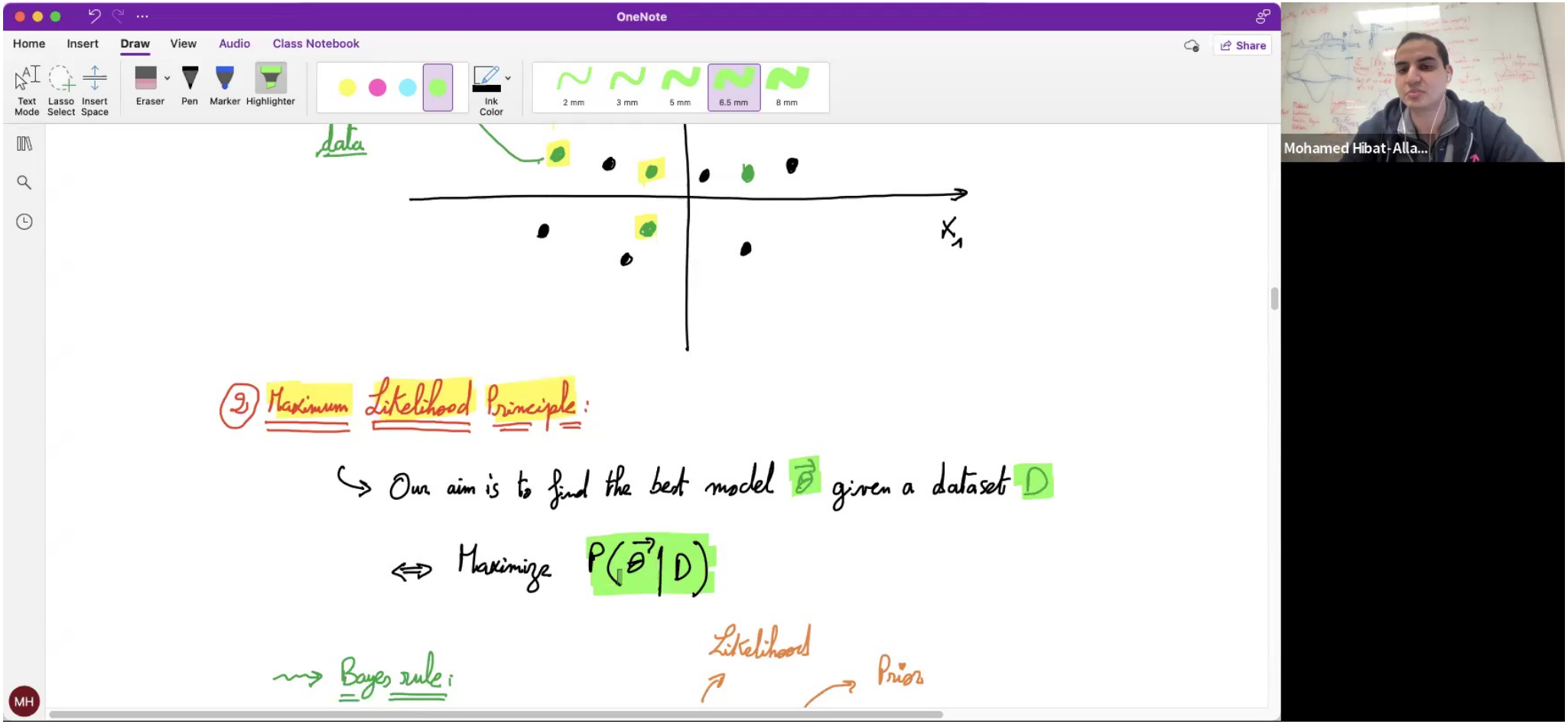

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
can also download the course materials to work through the material at your own pace.
Credit: https://app.inferkit.com/demo
* The goal:
↳ We have input training data from some unknown distribution « P_{data} » and we want to have a model so that $P_{model} \simeq P_{data}$.
D_{train}
Model
Training
$P_{model} \simeq P_{data}$
Sample new data points $\vec{x}$
$(D_{train} \simeq P_{data})$
✓ ?
MH
Mohamed Hibat-Alla...


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
D_{train}
Model
Training
$P_{model} \simeq P_{data}$
$(D_{train} \sim P_{data})$
?
Sample new data points $\vec{x}$
Training data
Generated data
x_2
x_1
MH
Mohamed Hibat-Alla...

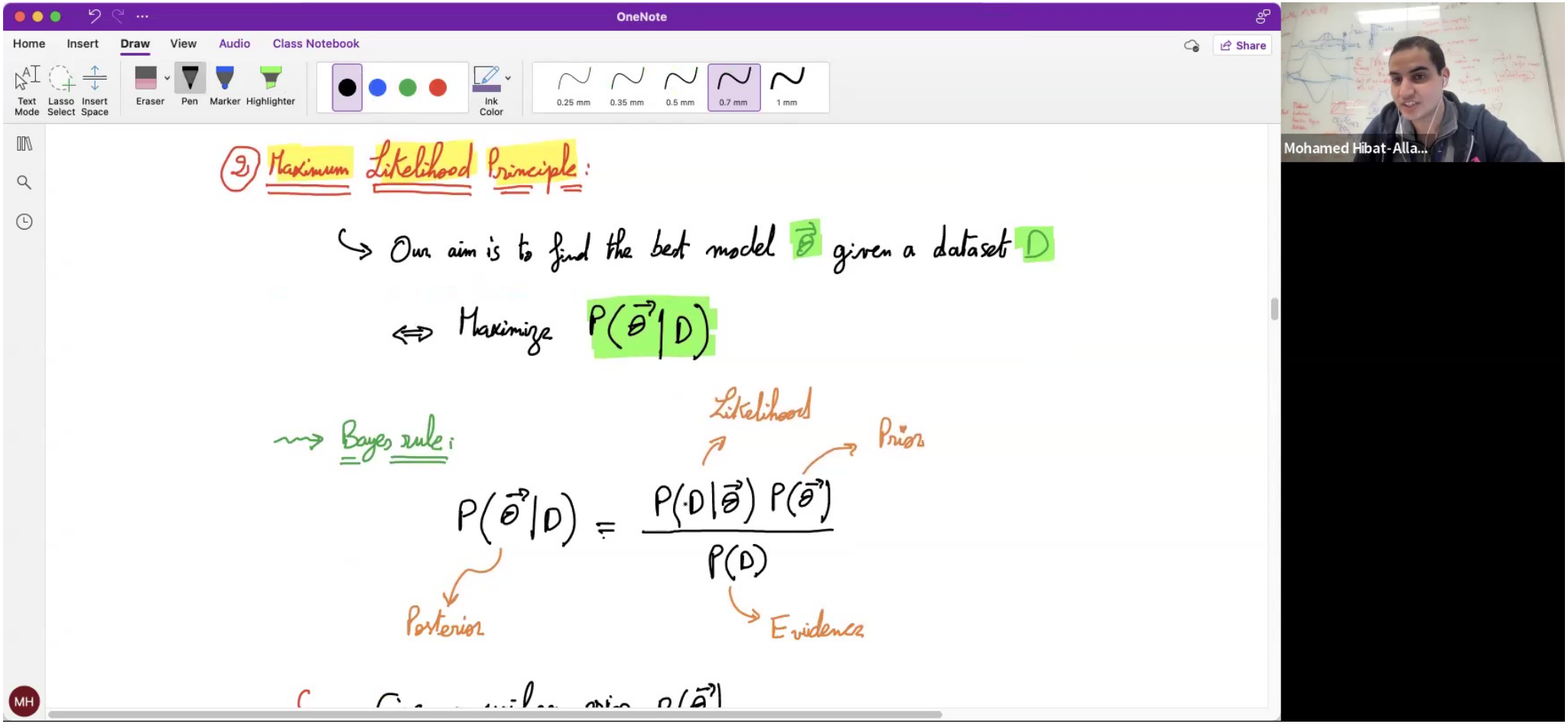

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
$(D_{train} \sim P_{data})$
✓ ?
Sample new data points $\vec{X}$
Training data
Generated data
X_2
X_1
(2) Maximum Likelihood Principle:
MH
Mohamed Hibat-Alla...

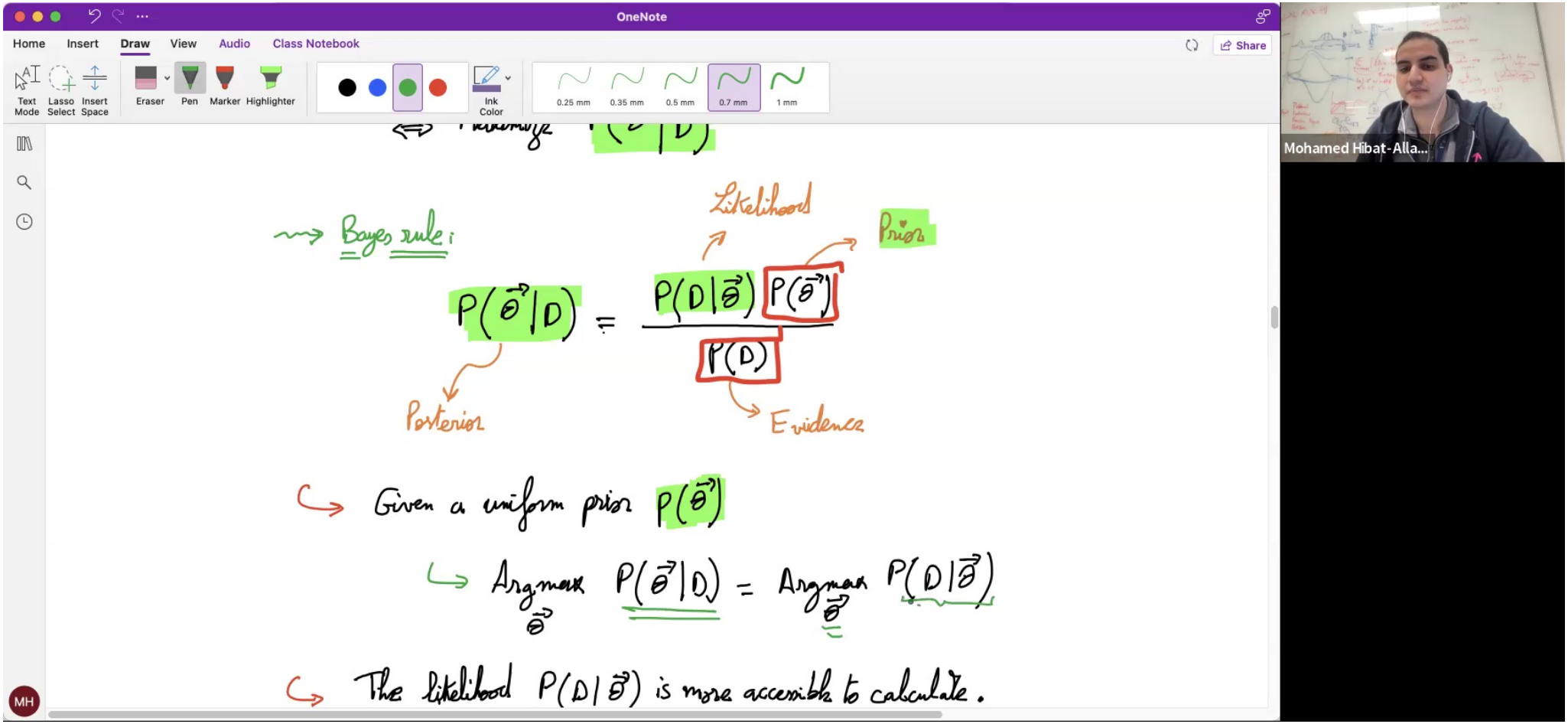

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Model
Training
Sample new data points
Training data
Generated data
Perimeter-A
MH

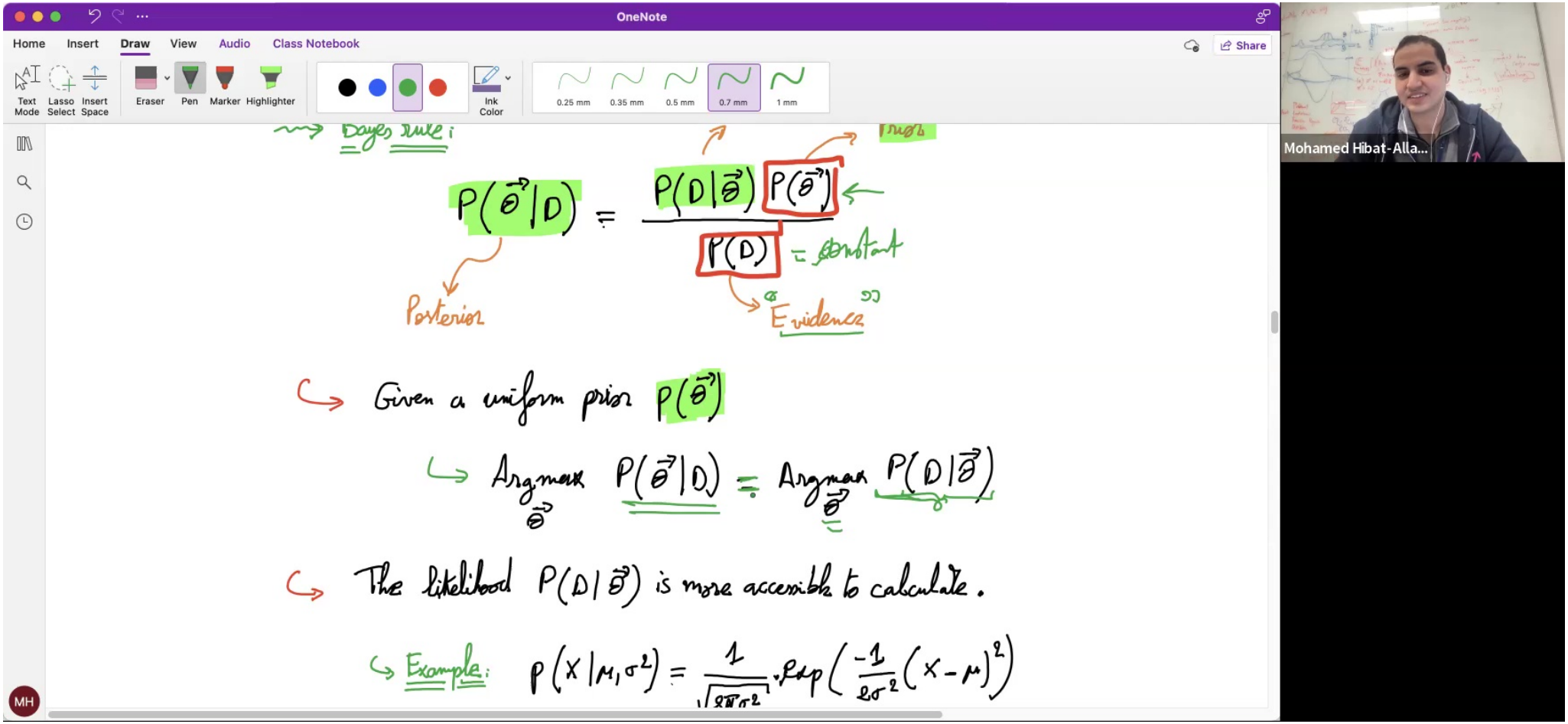

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Perimeter-A
" $p_{face}(\vec{x})$ "
$\vec{x}$ = image
Credit: this-person-does-not-exist.com
MH


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
data
X_1
② Maximum Likelihood Principle :
↳ Our aim is to find the best model $\vec{\theta}$ given a dataset D
⇔ Maximize $P(\vec{\theta}|D)$
⇝ Bayes rule :
Likelihood
Prior
MH

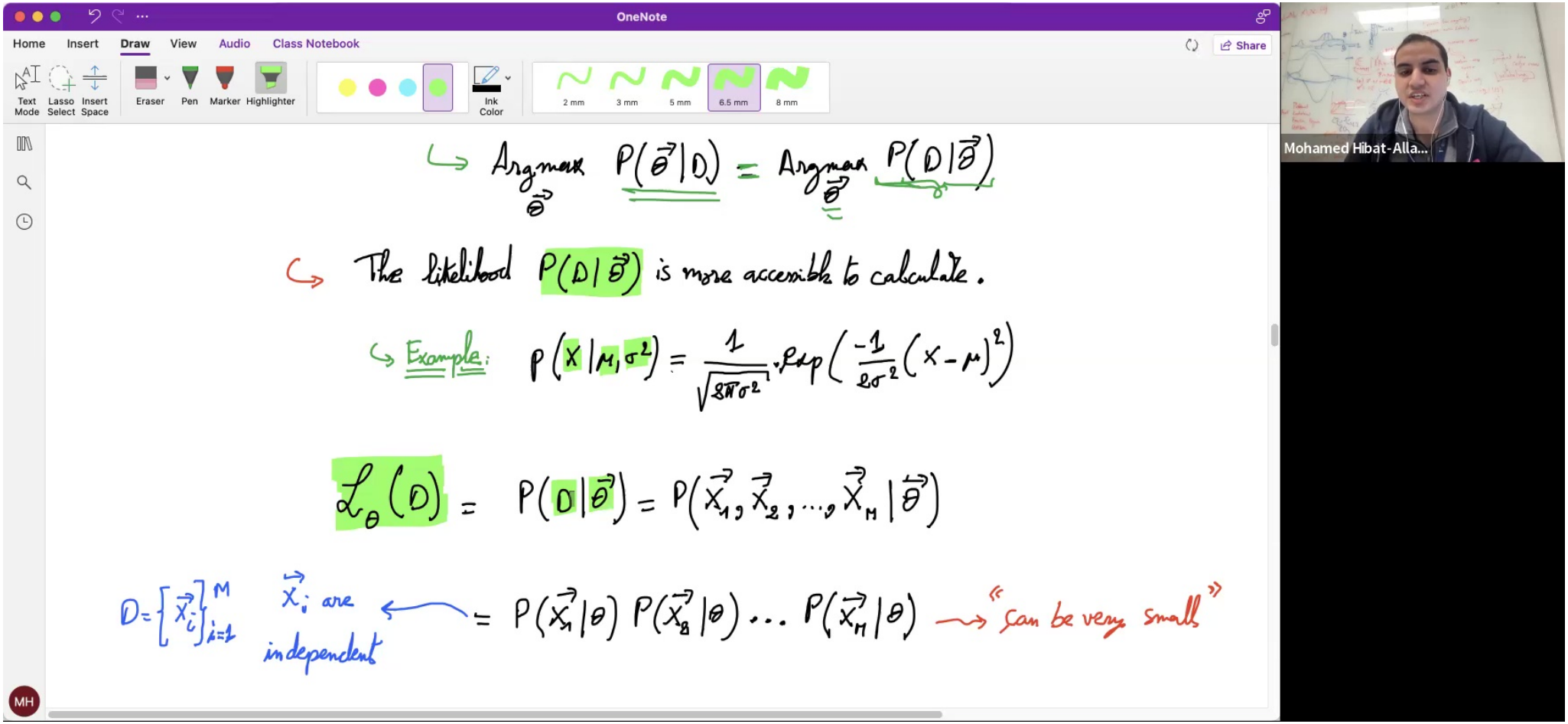
Mohamed Hibat-Alla...

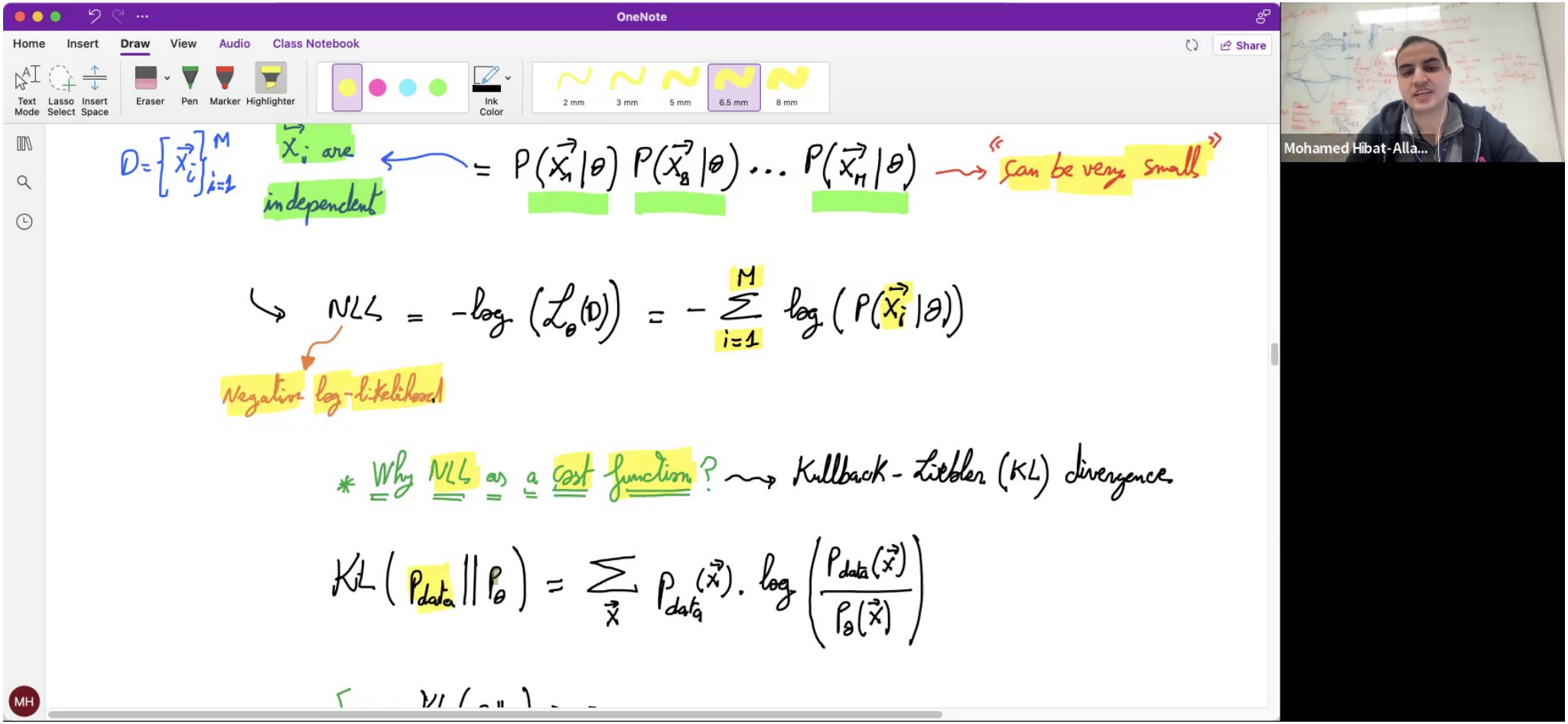

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Mohamed Hibat-Alla...
2) Maximum Likelihood Principle :
↳ Our aim is to find the best model $\vec{\theta}$ given a dataset D
⇔ Maximize $P(\vec{\theta}|D)$
⇝ Bayes rule :
Likelihood
Prior
$P(\vec{\theta}|D) = \frac{P(D|\vec{\theta})\,P(\vec{\theta})}{P(D)}$
Posterior
Evidence
MH

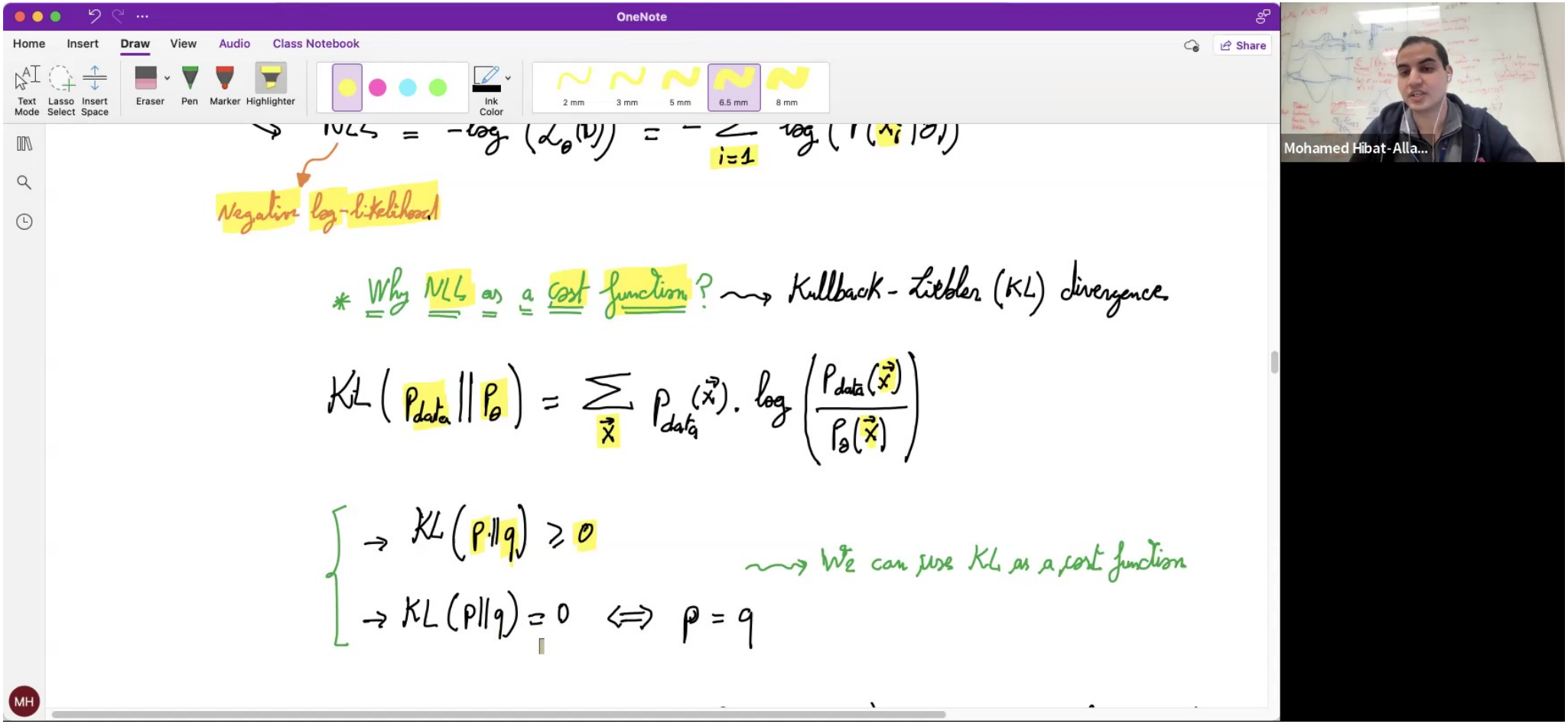

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Mohamed Hibat-Alla...
Bayes rule :
Likelihood
Prior
$P(\vec{\theta}|D) = \frac{P(D|\vec{\theta})\,P(\vec{\theta})}{P(D)}$
Posterior
Evidence
Given a uniform prior $P(\vec{\theta})$
$\text{Argmax}_{\vec{\theta}}\, P(\vec{\theta}|D) = \text{Argmax}_{\vec{\theta}}\, P(D|\vec{\theta})$
The likelihood $P(D|\vec{\theta})$ is more accessible to calculate.
MH


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Bayes rule:
$P(\vec{\theta}|D) = \frac{P(D|\vec{\theta})\,P(\vec{\theta})}{P(D)}$
Prior
Posterior
Evidence
= constant
Given a uniform prior $P(\vec{\theta})$
$\text{Argmax}_{\vec{\theta}}\ P(\vec{\theta}|D) = \text{Argmax}_{\vec{\theta}}\ P(D|\vec{\theta})$
The likelihood $P(D|\vec{\theta})$ is more accessible to calculate.
Example: $P(X|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-1}{2\sigma^2}(X-\mu)^2\right)$
MH

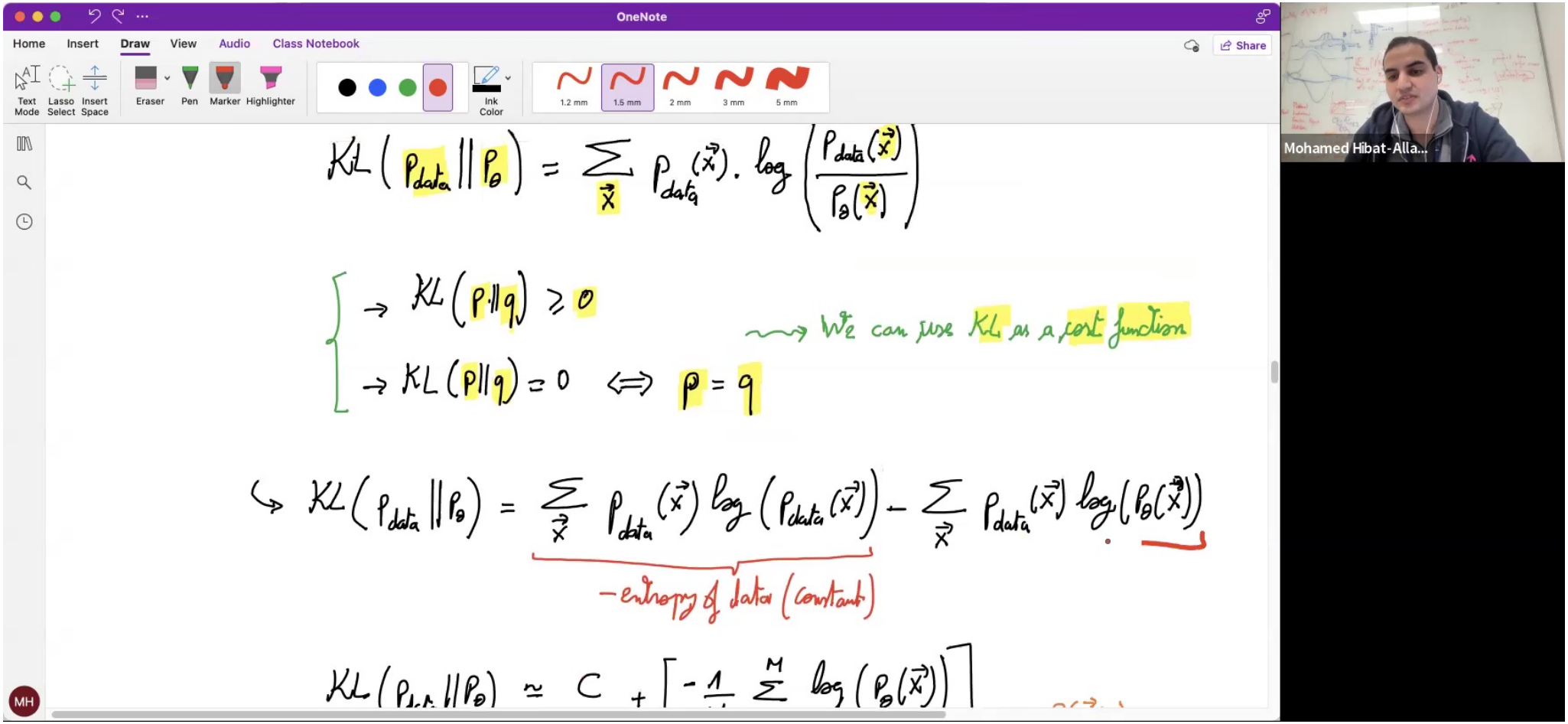
Mohamed Hibat-Alla...


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
$\hookrightarrow \underset{\vec{\theta}}{\text{Argmax}}\, P(\vec{\theta}|D) = \underset{\vec{\theta}}{\text{Argmax}}\, P(D|\vec{\theta})$
$\hookrightarrow$ The likelihood $P(D|\vec{\theta})$ is more accessible to calculate.
$\hookrightarrow$ Example: $P(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$
$\mathcal{L}_\theta(D) = P(D|\vec{\theta}) = P(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M|\vec{\theta})$
$D = \{\vec{x}_i\}_{i=1}^{M}$
$\vec{x}_i$ are independent
$= P(\vec{x}_1|\theta)\, P(\vec{x}_2|\theta) \dots P(\vec{x}_M|\theta)$
"can be very small"
MH
Mohamed Hibat-Alla...

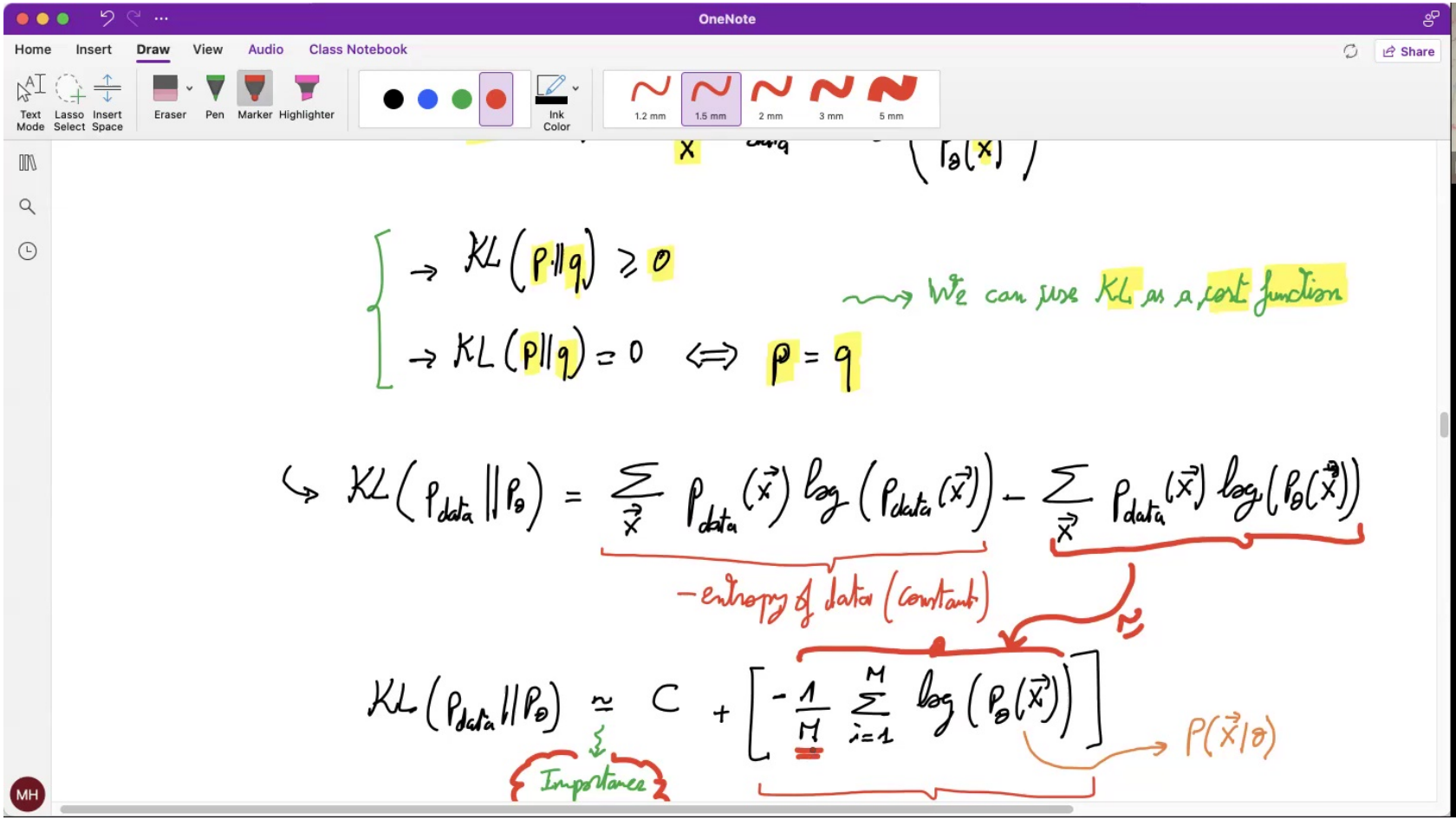

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
$D = \{\vec{x}_i\}_{i=1}^{M}$
$\vec{x}_i$ are independent
$= P(\vec{x}_1|\theta)\, P(\vec{x}_2|\theta) \dots P(\vec{x}_M|\theta)$
« can be very small »
$\hookrightarrow NLL = -\log(\mathcal{L}_\theta(D)) = -\sum_{i=1}^{M} \log(P(\vec{x}_i|\theta))$
Negative log-likelihood
* Why NLL as a cost function ? ⇝ Kullback-Leibler (KL) divergence
$KL(P_{data} \| P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \cdot \log\left(\frac{P_{data}(\vec{x})}{P_\theta(\vec{x})}\right)$
MH
Mohamed Hibat-Alla...

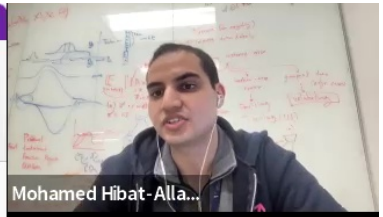

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
$\mathrm{NLL} = -\log(\mathcal{L}_\theta(\mathcal{D})) = -\sum_{i=1} \log(P(x_i|\theta))$
Negative log-likelihood
* Why NLL as a cost function ? ⇝ Kullback-Leibler (KL) divergence
$KL(P_{data} \| P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \cdot \log\left(\frac{P_{data}(\vec{x})}{P_\theta(\vec{x})}\right)$
→ $KL(p\|q) \geq 0$
→ $KL(p\|q) = 0 \iff p = q$
⇝ We can use KL as a cost function
MH

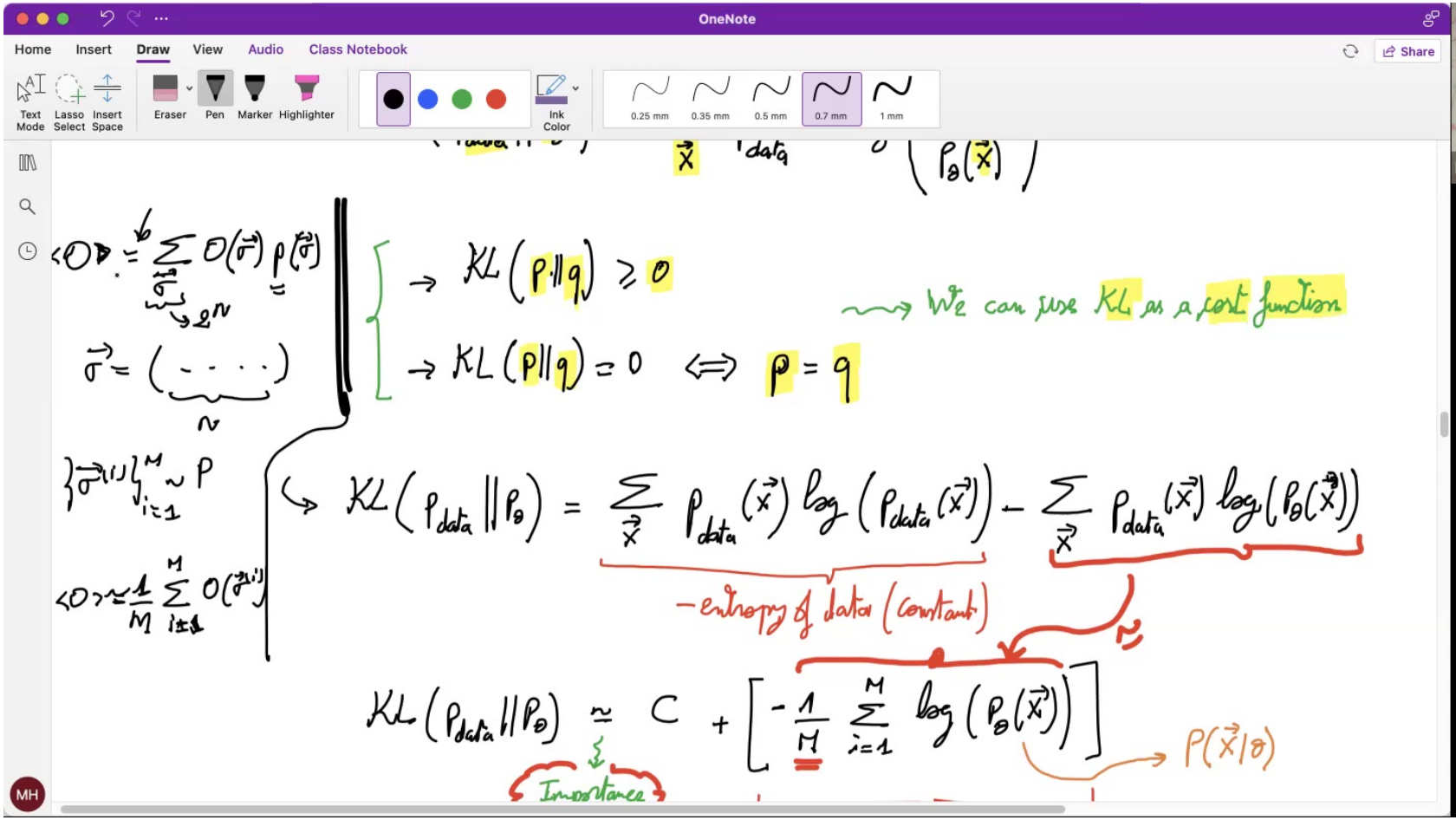
Mohamed Hibat-Alla...

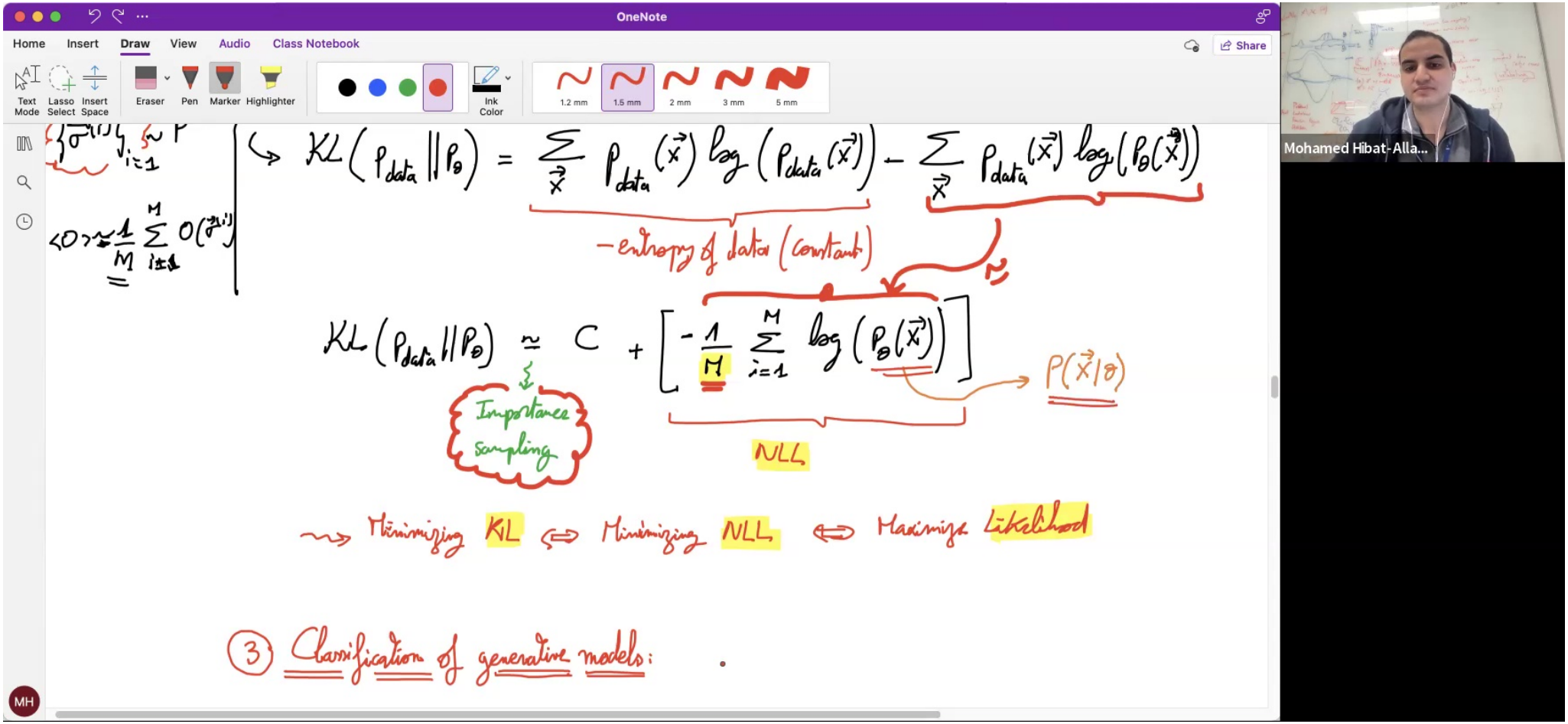

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$KL(P_{data} \| P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \cdot \log\left(\frac{P_{data}(\vec{x})}{P_\theta(\vec{x})}\right)$
$\rightarrow KL(P\|q) \geq 0$
$\rightarrow KL(P\|q) = 0 \iff P = q$
We can use KL as a cost function
$\hookrightarrow KL(P_{data} \| P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \log\left(P_{data}(\vec{x})\right) - \sum_{\vec{x}} P_{data}(\vec{x}) \log\left(P_\theta(\vec{x})\right)$
- entropy of data (constant)
$KL(P_{data} \| P_\theta) \simeq C + \left[-\frac{1}{M} \sum^{M} \log\left(P_\theta(\vec{x})\right)\right]$
MH

Mohamed Hibat-Alla...

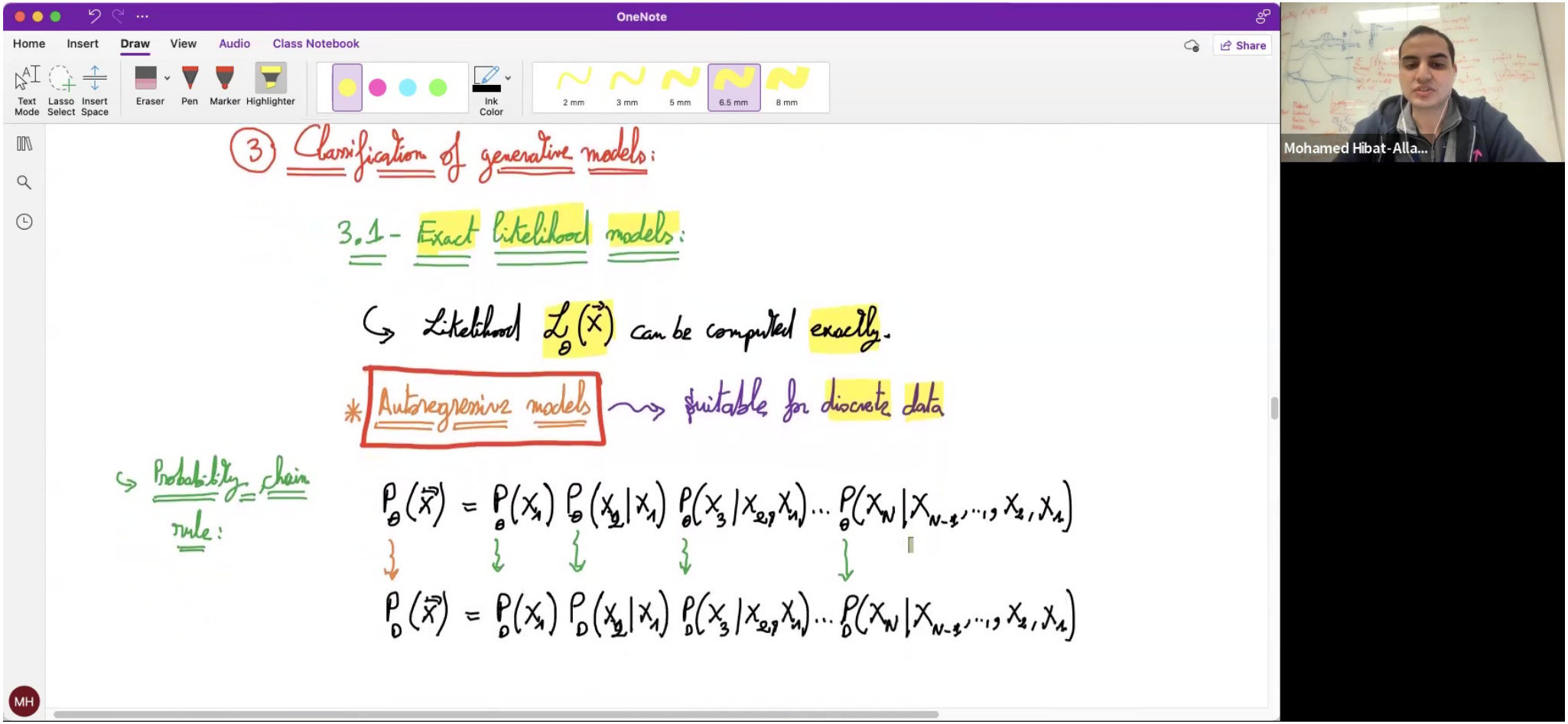

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$\rightarrow KL(p\|q) \geq 0$
$\rightarrow KL(p\|q) = 0 \iff p = q$
We can use KL as a cost function
$\hookrightarrow KL(p_{data}\|p_\theta) = \sum_{\vec{x}} p_{data}(\vec{x}) \log(p_{data}(\vec{x})) - \sum_{\vec{x}} p_{data}(\vec{x}) \log(p_\theta(\vec{x}))$
- entropy of data (constant)
$KL(p_{data}\|p_\theta) \simeq C + \left[-\frac{1}{M}\sum_{i=1}^{M} \log(p_\theta(\vec{x}))\right]$
Importance
$P(\vec{x}|\theta)$
MH


Mohamed Hibat-Alla...

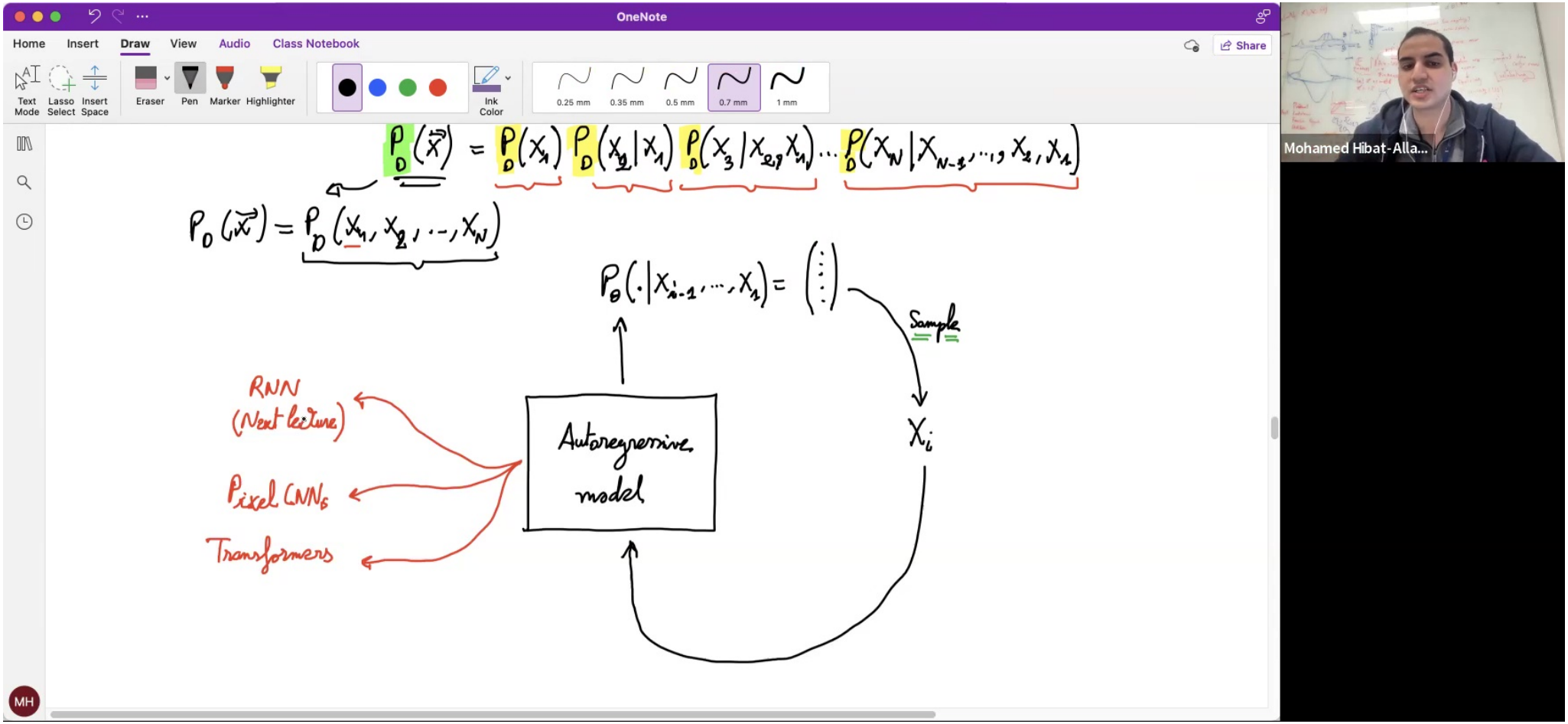

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
$\langle O \rangle = \sum_{\vec{\sigma}} O(\vec{\sigma})\, p(\vec{\sigma})$
$\vec{\sigma} = (\cdots)$
N
$\{\vec{\sigma}^{(i)}\}_{i=1}^{M} \sim p$
$\langle O \rangle \simeq \frac{1}{M} \sum_{i=1}^{M} O(\vec{\sigma}^{(i)})$
$KL(p\|q) \geq 0$
$KL(p\|q) = 0 \iff p = q$
We can use KL as a cost function
$KL(p_{data}\|p_\theta) = \sum_{\vec{x}} p_{data}(\vec{x}) \log\left(p_{data}(\vec{x})\right) - \sum_{\vec{x}} p_{data}(\vec{x}) \log\left(p_\theta(\vec{x})\right)$
- entropy of data (constant)
$KL(p_{data}\|p_\theta) \simeq C + \left[-\frac{1}{M} \sum_{i=1}^{M} \log\left(p_\theta(\vec{x})\right)\right]$
Importance
$P(\vec{x}|\theta)$
Mohamed Hibat-Alla...
MH


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$KL(P_{data} \| P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \log\left(P_{data}(\vec{x})\right) - \sum_{\vec{x}} P_{data}(\vec{x}) \log\left(P_\theta(\vec{x})\right)$
- entropy of data (constant)
$KL(P_{data} \| P_\theta) \simeq C + \left[-\frac{1}{M} \sum_{i=1}^{M} \log\left(P_\theta(\vec{x})\right)\right]$
$P(\vec{x}|\theta)$
Importance sampling
NLL
Minimizing KL ⇔ Minimizing NLL ⇔ Maximize Likelihood
3) Classification of generative models:
MH

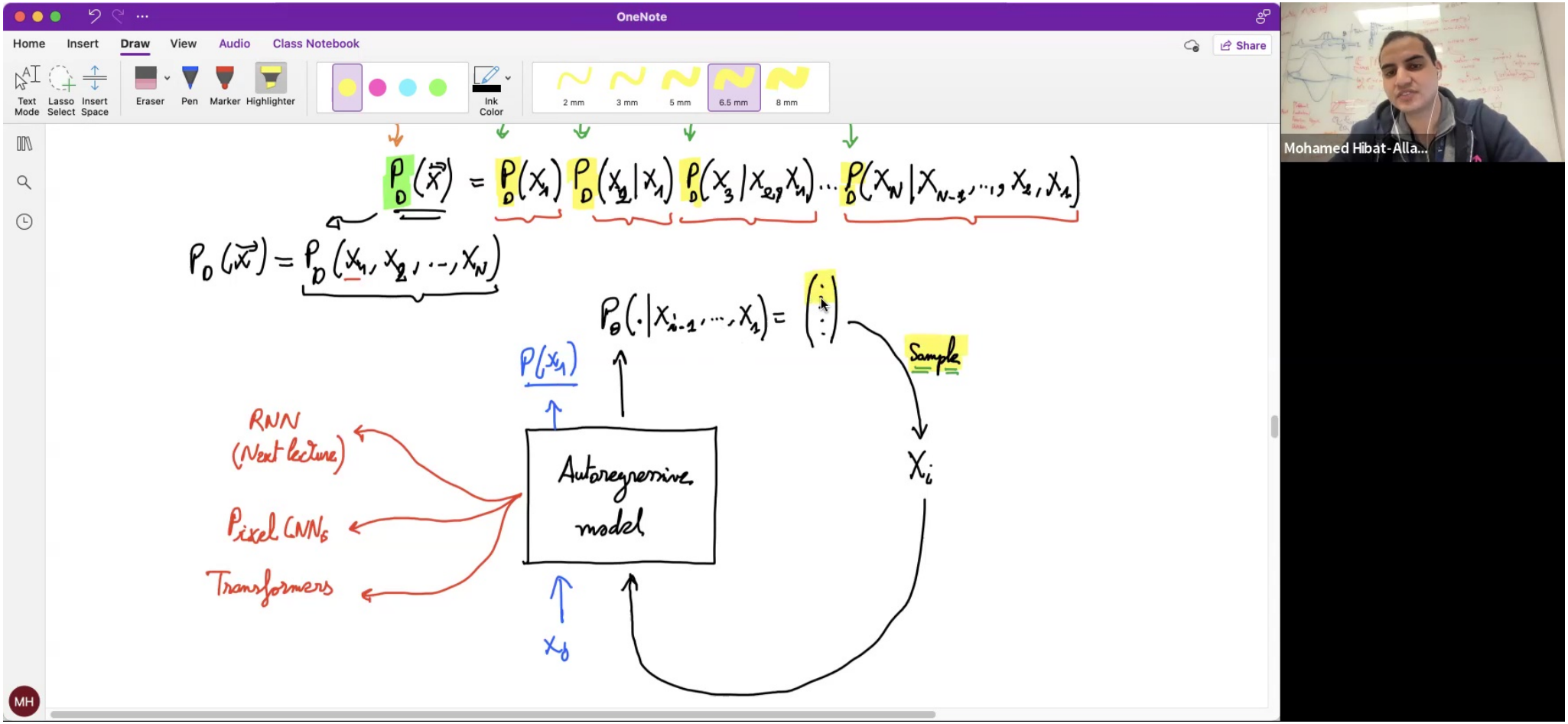
Mohamed Hibat-Alla...

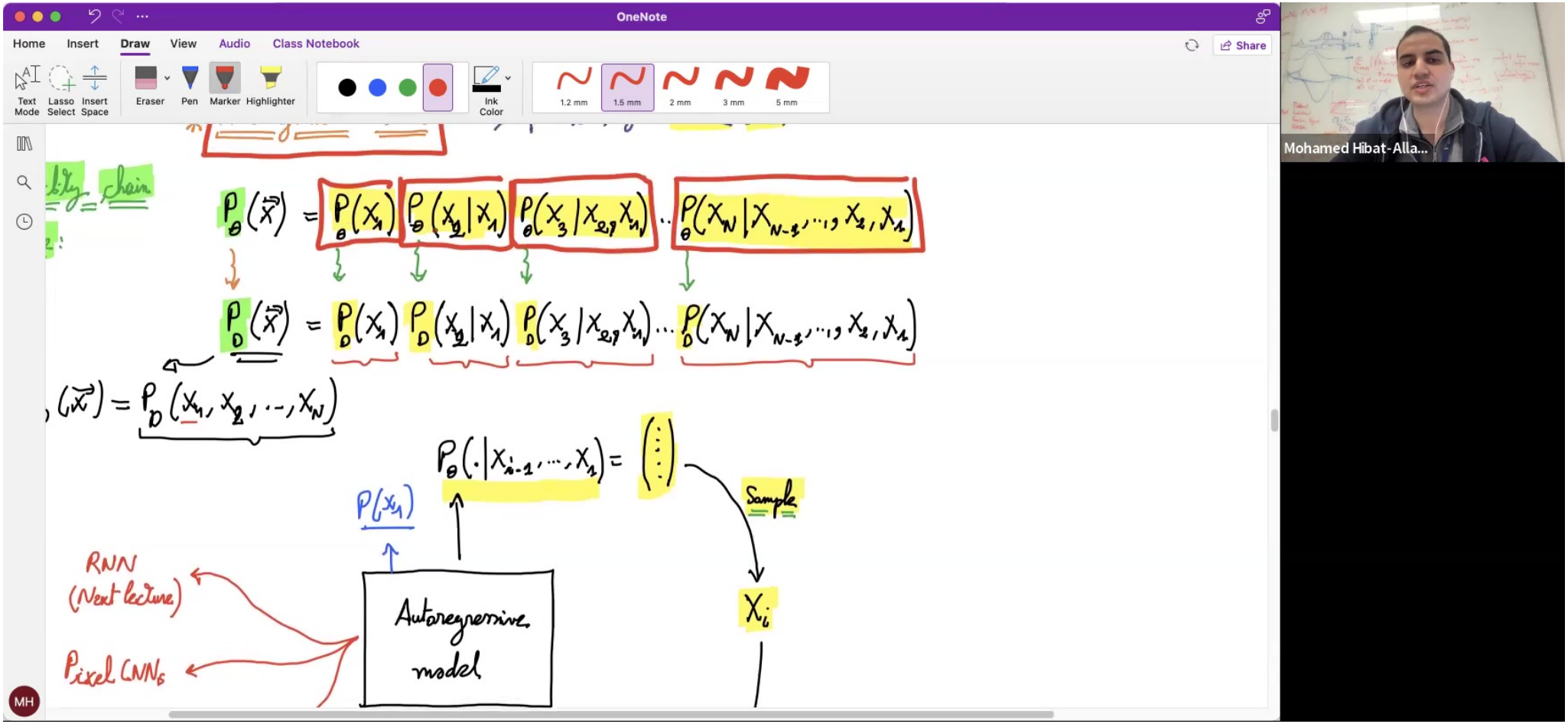

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
3) Classification of generative models:
3.1- Exact likelihood models:
↳ Likelihood $\mathcal{L}_\theta(\vec{x})$ can be computed exactly.
* Autoregressive models ⇝ suitable for discrete data
↳ Probability chain rule:
$P_\theta(\vec{x}) = P_\theta(x_1)\, P_\theta(x_2|x_1)\, P_\theta(x_3|x_2,x_1) \dots P_\theta(x_N|x_{N-1},\dots,x_2,x_1)$
$P_D(\vec{x}) = P_D(x_1)\, P_D(x_2|x_1)\, P_D(x_3|x_2,x_1) \dots P_D(x_N|x_{N-1},\dots,x_2,x_1)$
MH

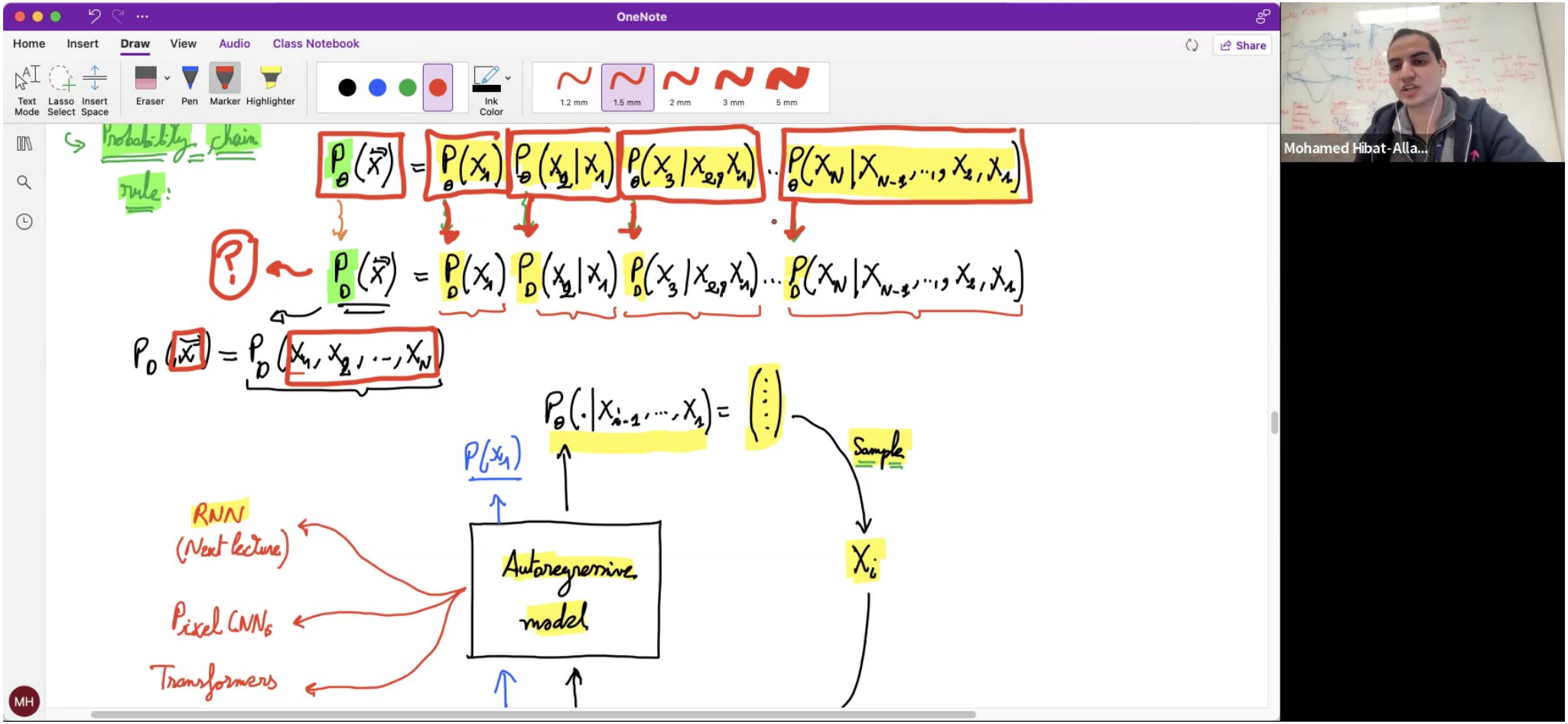
Mohamed Hibat-Alla...

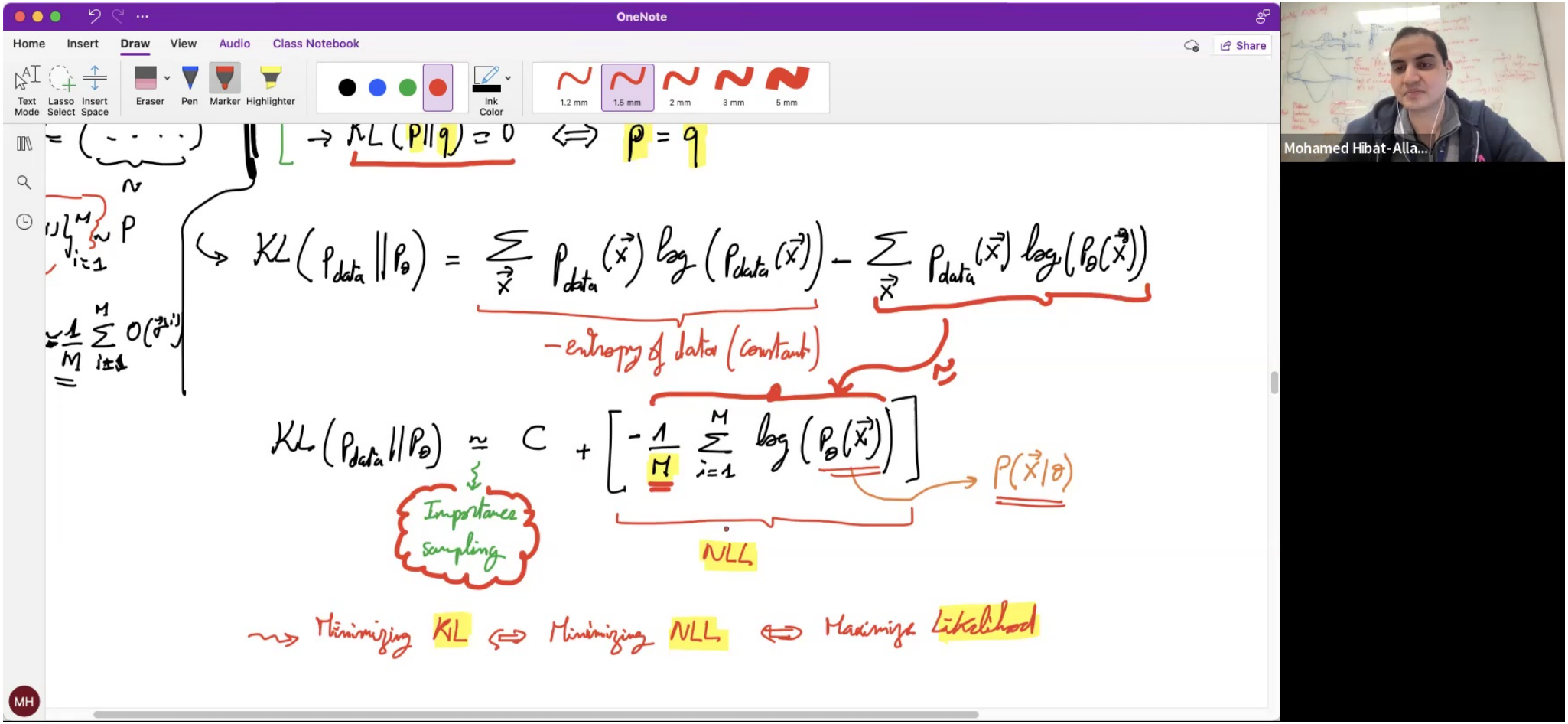

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
$p_D(\vec{x}) = p_D(x_1)\, p_D(x_2|x_1)\, p_D(x_3|x_2,x_1) \cdots p_D(x_N|x_{N-1},\dots,x_2,x_1)$
$p_D(\vec{x}) = p_D(x_1, x_2, \dots, x_N)$
$p_\theta(\cdot|x_{i-1},\dots,x_1) = \begin{pmatrix}\vdots\end{pmatrix}$
Sample
x_i
Autoregressive model
RNN (Next lecture)
PixelCNNs
Transformers
MH

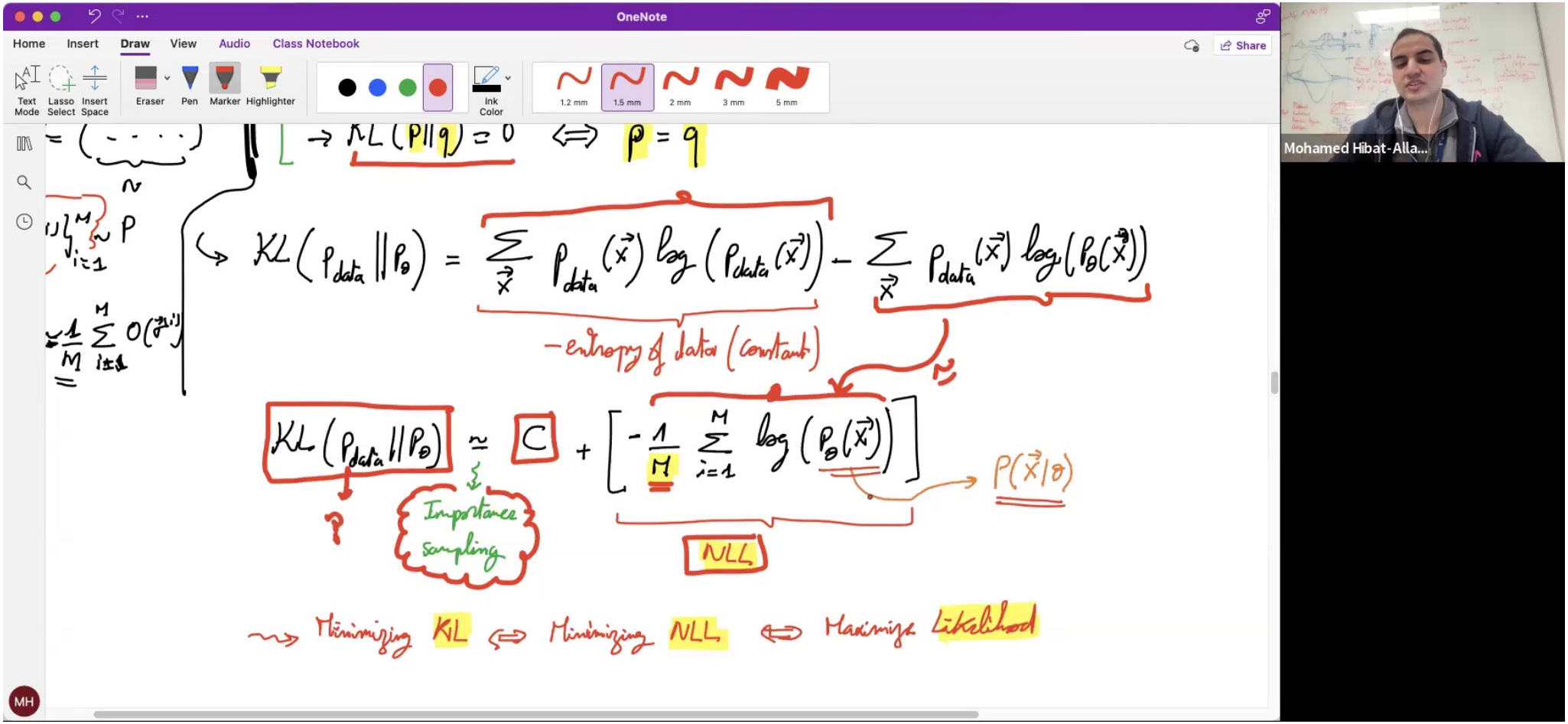
Mohamed Hibat-Alla...

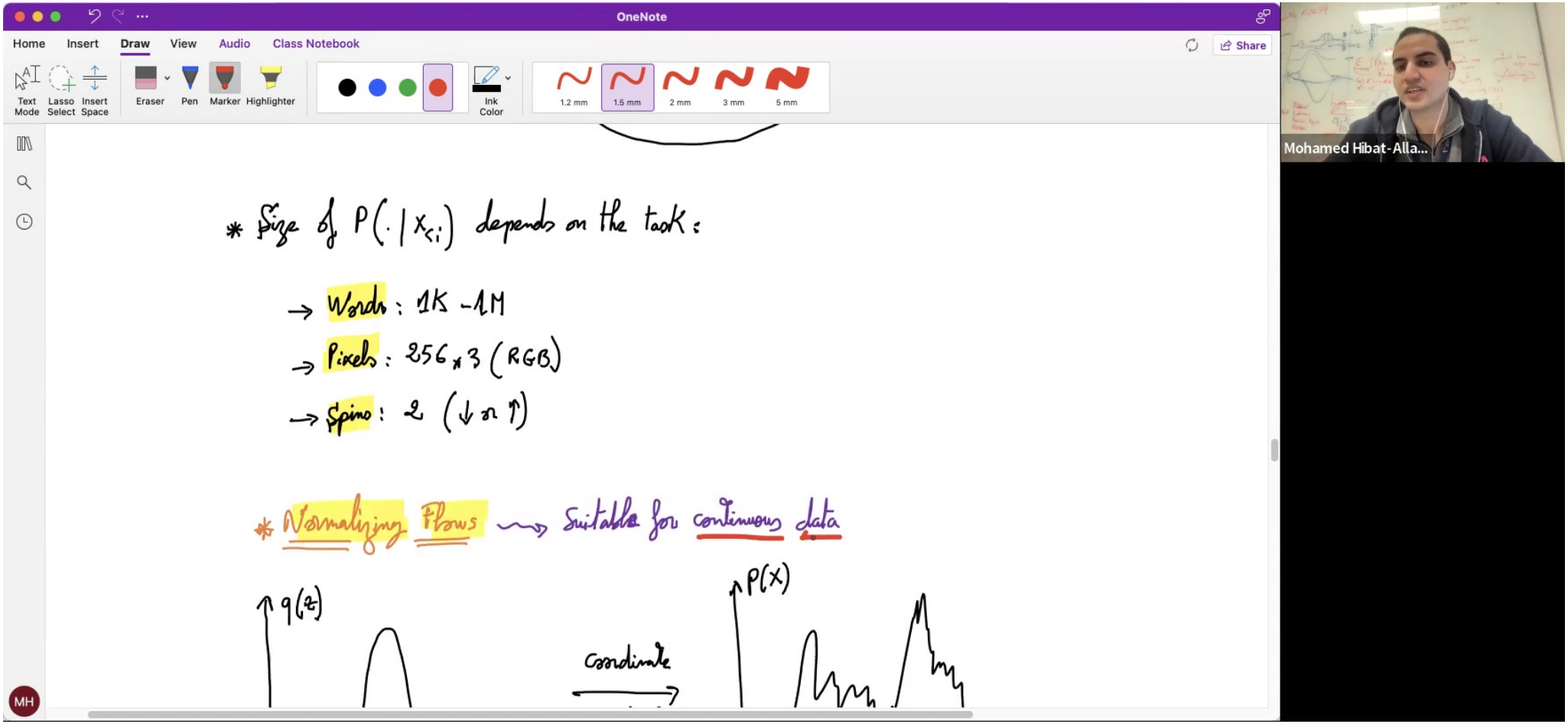

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
$p_D(\vec{x}) = p_D(x_1)\, p_D(x_2|x_1)\, p_D(x_3|x_2,x_1) \cdots p_D(x_N|x_{N-1},\dots,x_2,x_1)$
$p_\theta(\vec{x}) = p_\theta(x_1, x_2, \dots, x_N)$
$p_\theta(\cdot|x_{i-1},\dots,x_1) =$
Sample
$p(x_1)$
RNN
(Next lecture)
PixelCNNs
Transformers
Autoregressive model
x_i
x_0
Mohamed Hibat-Alla...
MH


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
Mohamed Hibat-Alla...
$p_\theta(\vec{x}) = p_\theta(x_1)\, p_\theta(x_2|x_1)\, p_\theta(x_3|x_2,x_1) \cdots p_\theta(x_N|x_{N-1},\dots,x_2,x_1)$
$p_D(\vec{x}) = p_D(x_1)\, p_D(x_2|x_1)\, p_D(x_3|x_2,x_1) \cdots p_D(x_N|x_{N-1},\dots,x_2,x_1)$
$p_D(\vec{x}) = p_D(x_1, x_2, \dots, x_N)$
$p_\theta(\cdot|x_{i-1},\dots,x_1) = (\dots)$
Sample
x_i
$p(x_1)$
RNN
(Next lecture)
PixelCNNs
Autoregressive model
MH

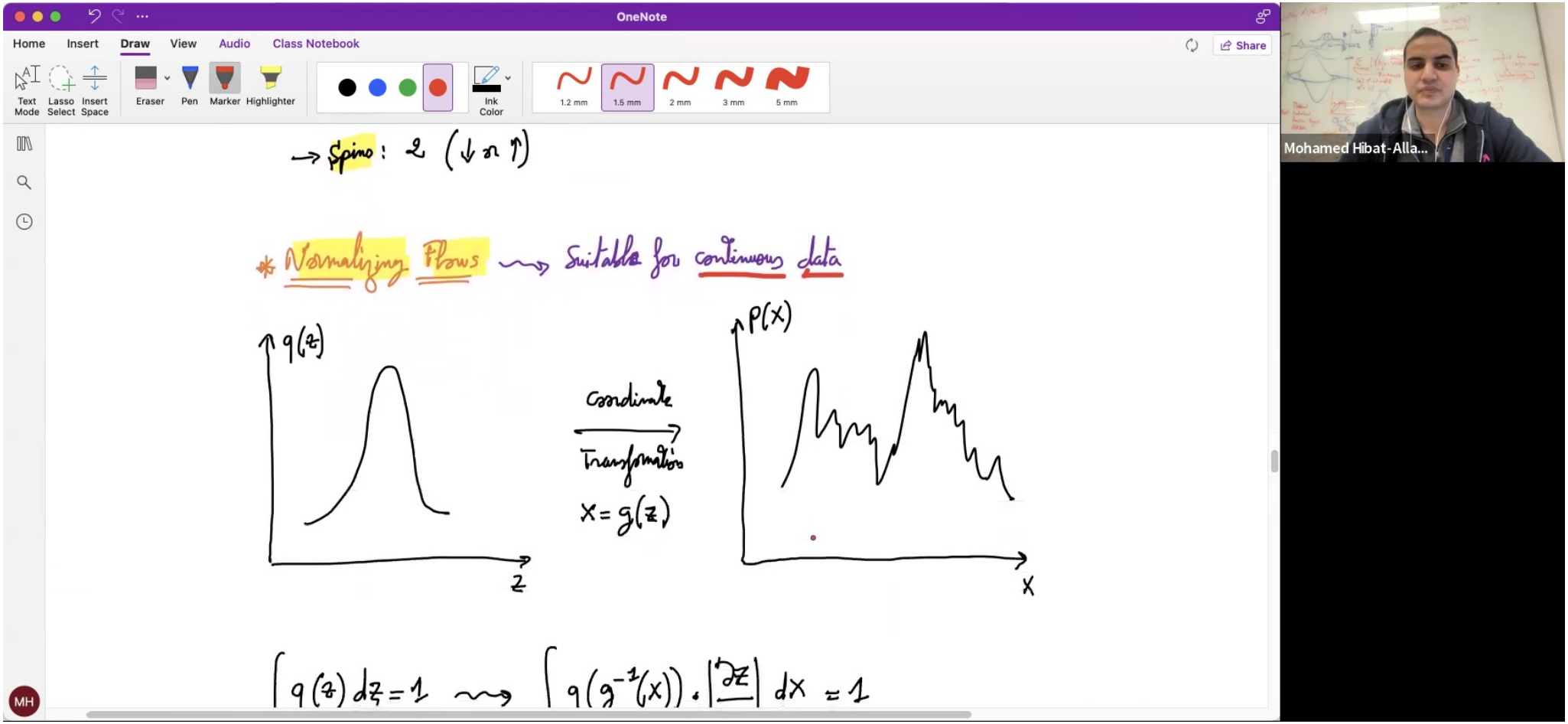

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
Probability chain rule:
RNN
(Next lecture)
PixelCNNs
Transformers
Autoregressive model
Sample
MH
Mohamed Hibat-Alla...


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$KL(P\|q) \geq 0 \Leftrightarrow P = q$
$KL(P_{data}\|P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \log(P_{data}(\vec{x})) - \sum_{\vec{x}} P_{data}(\vec{x}) \log(P_\theta(\vec{x}))$
-entropy of data (constant)
$KL(P_{data}\|P_\theta) \simeq C + \left[-\frac{1}{M}\sum_{i=1}^{M} \log(P_\theta(\vec{x}^{(i)}))\right]$
Importance sampling
NLL
$P(\vec{x}|\theta)$
Minimizing KL $\Leftrightarrow$ Minimizing NLL $\Leftrightarrow$ Maximize Likelihood
Mohamed Hibat-Alla...
MH

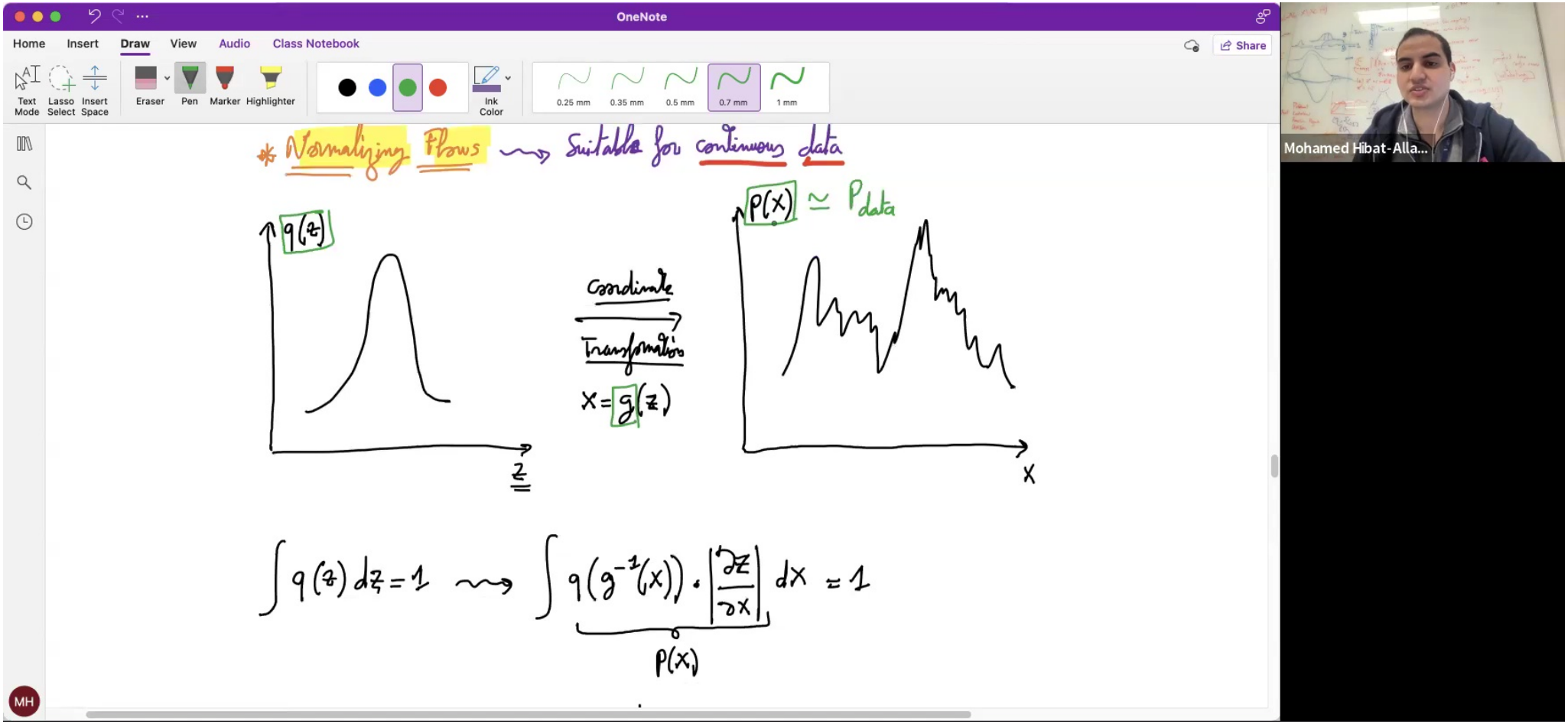

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$KL(P\|q) \geq 0 \Leftrightarrow P = q$
$KL(P_{data}\|P_\theta) = \sum_{\vec{x}} P_{data}(\vec{x}) \log(P_{data}(\vec{x})) - \sum_{\vec{x}} P_{data}(\vec{x}) \log(P_\theta(\vec{x}))$
- entropy of data (constant)
$KL(P_{data}\|P_\theta) \simeq C + \left[-\frac{1}{M}\sum_{i=1}^{M} \log(P_\theta(\vec{x}))\right]$
Importance sampling
NLL
$P(\vec{x}|\theta)$
Minimizing KL $\Leftrightarrow$ Minimizing NLL $\Leftrightarrow$ Maximize Likelihood
MH
Mohamed Hibat-Alla...

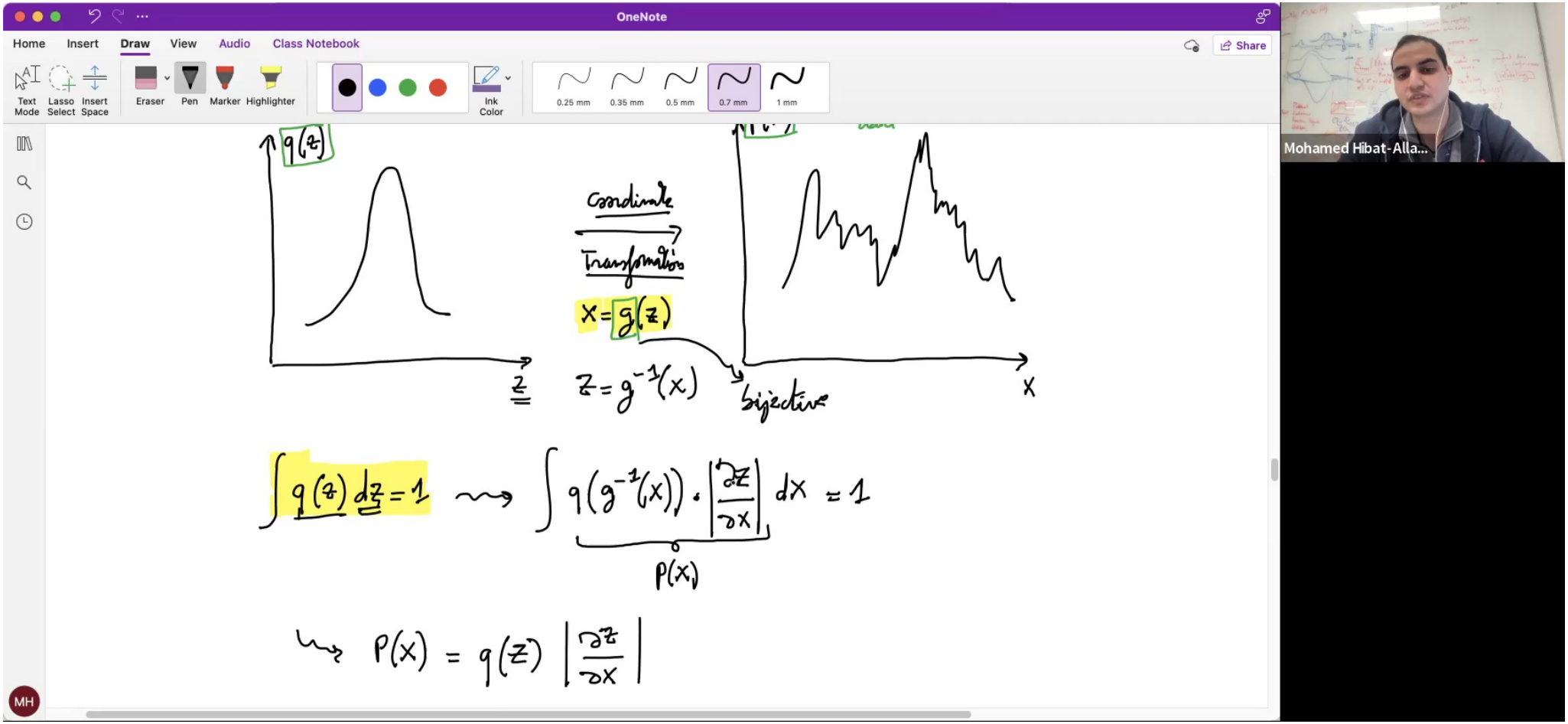

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
* Size of $P(\cdot | x_{<i})$ depends on the task:
→ Words: 1K - 1M
→ Pixels: 256 x 3 (RGB)
→ Spins: 2 (↓ or ↑)
* Normalizing Flows ⇝ Suitable for continuous data
q(z)
Coordinate
P(x)
MH

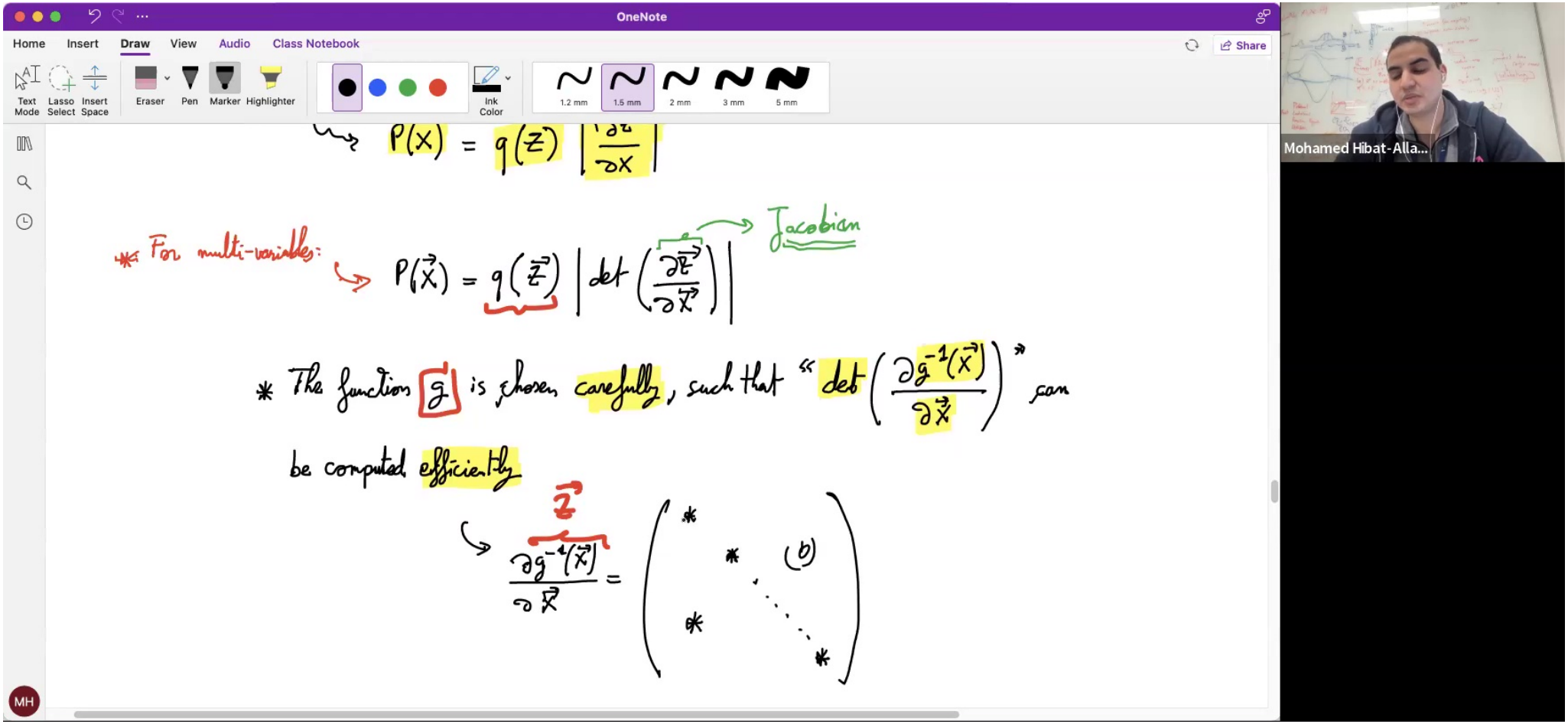
Mohamed Hibat-Alla...


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
→ Spins: 2 (↓ or ↑)
* Normalizing Flows → Suitable for continuous data
q(z)
z
Coordinate Transformation
X = g(z)
p(x)
x
∫ q(z) dz = 1 → ∫ q(g⁻¹(x)) · |∂z| dx = 1
MH

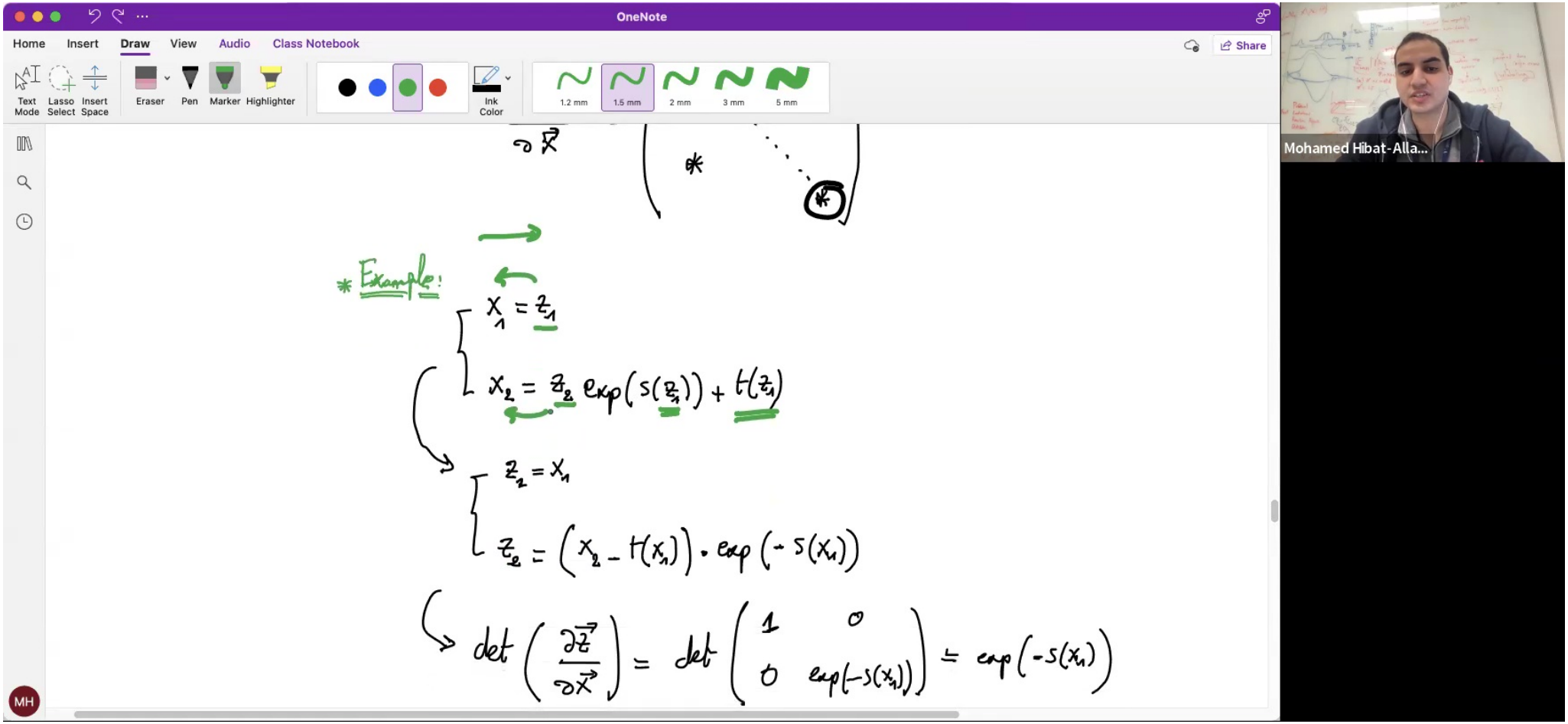
Mohamed Hibat-Alla...


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
* Normalizing Flows ⇝ Suitable for continuous data
q(z)
z
coordinate Transformation
x = g(z)
p(x) ≃ p_data
x
∫ q(z) dz = 1 ⇝ ∫ q(g⁻¹(x)) · |∂z/∂x| dx = 1
p(x)
MH
Mohamed Hibat-Alla...

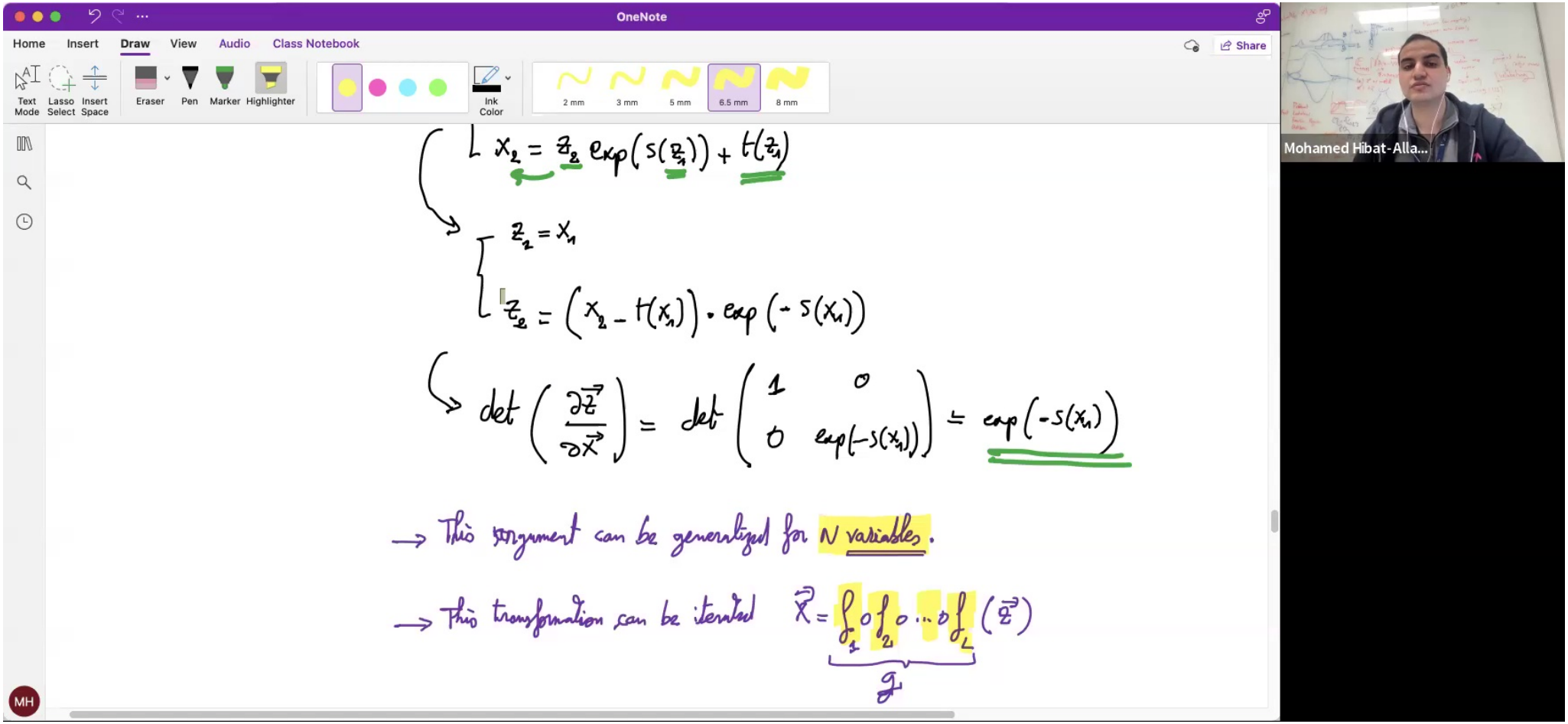

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
Mohamed Hibat-Alla...
MH
q(z)
z
Coordinate Transformation
X = g(z)
z = g⁻¹(X)
bijective
X
∫ q(z) dz = 1
∫ q(g⁻¹(X)) · |∂z/∂X| dX = 1
P(X)
P(X) = q(z) |∂z/∂X|


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$P(x) = q(z) \left| \frac{\partial z}{\partial x} \right|$
* For multi-variables:
$P(\vec{x}) = q(\vec{z}) \left| \det \left(\frac{\partial \vec{z}}{\partial \vec{x}} \right) \right|$
Jacobian
* The function g is chosen carefully, such that "$\det \left(\frac{\partial g^{-1}(\vec{x})}{\partial \vec{x}} \right)$" can be computed efficiently
$\frac{\partial g^{-1}(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} * & & & \\ & * & & (0) \\ & & \ddots & \\ & * & & * \end{pmatrix}$
MH

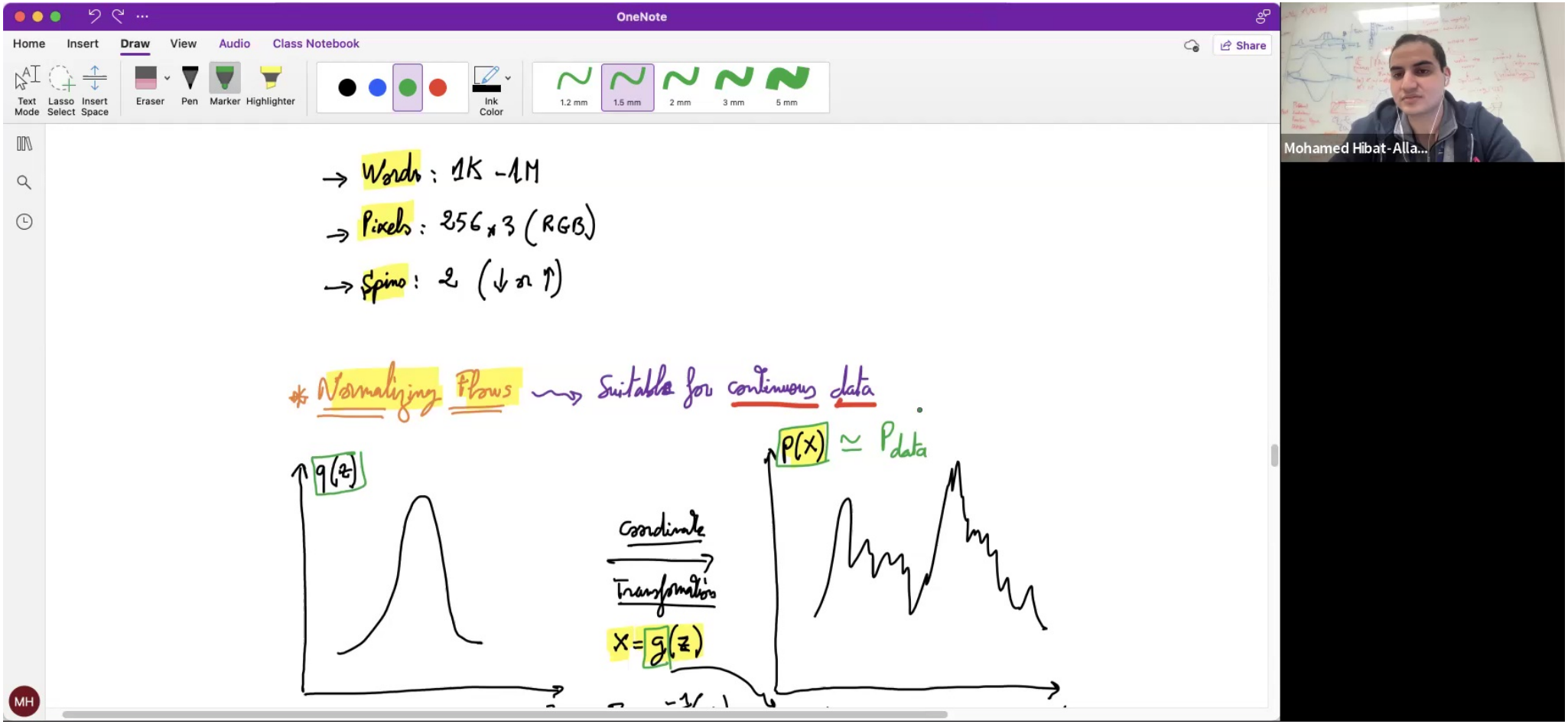
Mohamed Hibat-Alla...

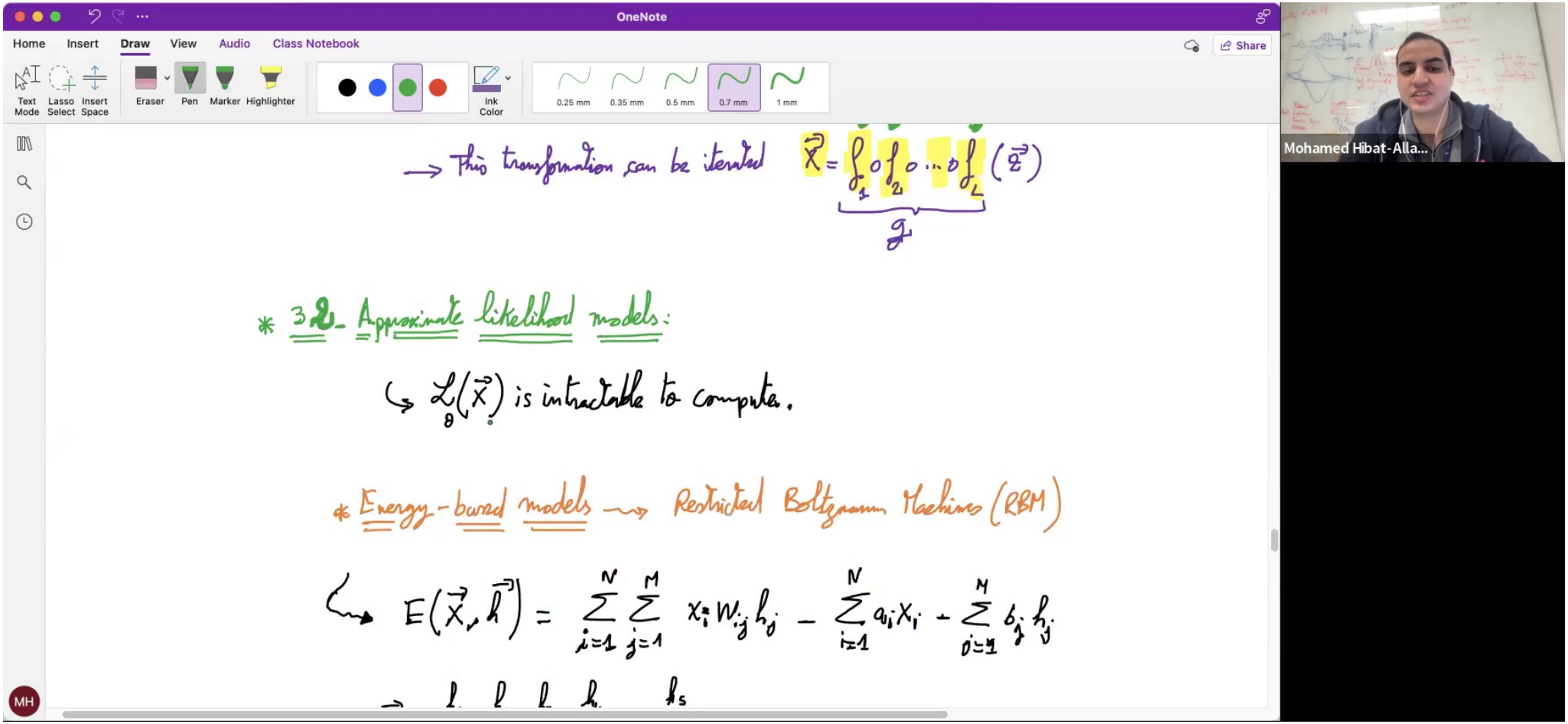

OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
Example:
$x_1 = z_1$
$x_2 = z_2 \exp(s(z_1)) + t(z_1)$
$z_2 = x_1$
$z_2 = (x_2 - t(x_1)) \cdot \exp(-s(x_1))$
$\det\left(\frac{\partial \vec{z}}{\partial \vec{x}}\right) = \det\begin{pmatrix} 1 & 0 \\ 0 & \exp(-s(x_1)) \end{pmatrix} = \exp(-s(x_1))$
MH

Mohamed Hibat-Alla...


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
$x_2 = z_2 \exp(s(z_1)) + t(z_1)$
$z_1 = x_1$
$z_2 = (x_2 - t(x_1)) \cdot \exp(-s(x_1))$
$\det\left(\frac{\partial \vec{z}}{\partial \vec{x}}\right) = \det\begin{pmatrix} 1 & 0 \\ 0 & \exp(-s(x_1)) \end{pmatrix} = \exp(-s(x_1))$
→ This argument can be generalized for N variables.
→ This transformation can be iterated $\vec{x} = f_1 \circ f_2 \circ \dots \circ f_L(\vec{z})$
g
MH

Mohamed Hibat-Alla...


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
→ Words : 1K - 1M
→ Pixels : 256 x 3 (RGB)
→ Spins : 2 (↓ or ↑)
* Normalizing Flows ⇝ Suitable for continuous data
q(z)
Coordinate Transformation
X = g(z)
p(x) ≃ p_data
MH
Mohamed Hibat-Alla...

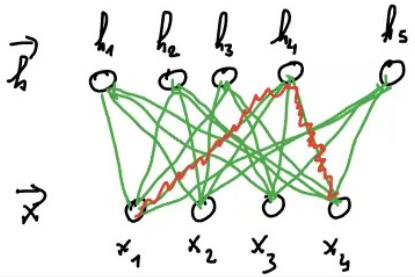

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
→ This transformation can be iterated $\vec{x} = \underbrace{f_1 \circ f_2 \circ \dots \circ f_L}_{g}(\vec{z})$
* 3.2- Approximate likelihood models:
↳ $\mathcal{L}_\theta(\vec{x})$ is intractable to compute.
* Energy-based models → Restricted Boltzmann Machines (RBM)
↳ $E(\vec{x}, \vec{h}) = \sum_{i=1}^{N}\sum_{j=1}^{M} x_i W_{ij} h_j - \sum_{i=1}^{N} a_i x_i - \sum_{j=1}^{M} b_j h_j$
→ h_1 h_2 h_3 h_4 h_5
MH
Mohamed Hibat-Alla...

* 3.2- Approximate likelihood models:

↳ $\mathcal{L}_\theta(\vec{x})$ is intractable to compute.

* Energy-based models ⇝ Restricted Boltzmann Machines (RBM)

↳ $E(\vec{x}, \vec{h}) = \sum_{i=1}^{N}\sum_{j=1}^{M} x_i w_{ij} h_j - \sum_{i=1}^{N} a_i x_i - \sum_{j=1}^{M} b_j h_j$



⇝ Correlation between x_1 and x_4 is introduced through h_4.

OneNote

Home Insert Draw View Audio Class Notebook

Share

Text Mode | Lasso Select | Insert Space | Eraser | Pen | Marker | Highlighter | Ink Color | 0.25 mm | 0.35 mm | 0.5 mm | 0.7 mm | 1 mm

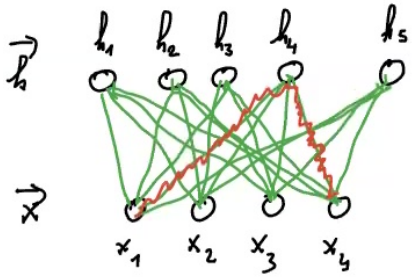
Mohamed Hibat-Alla...

* Energy-based models ⇝ Restricted Boltzmann Machines (RBM)

$$E(\vec{x}, \vec{h}) = \sum_{i=1}^{N} \sum_{j=1}^{M} x_i W_{ij} h_j - \sum_{i=1}^{N} a_i x_i - \sum_{j=1}^{M} b_j h_j$$

$\vec{x} = (\ ) \updownarrow N$

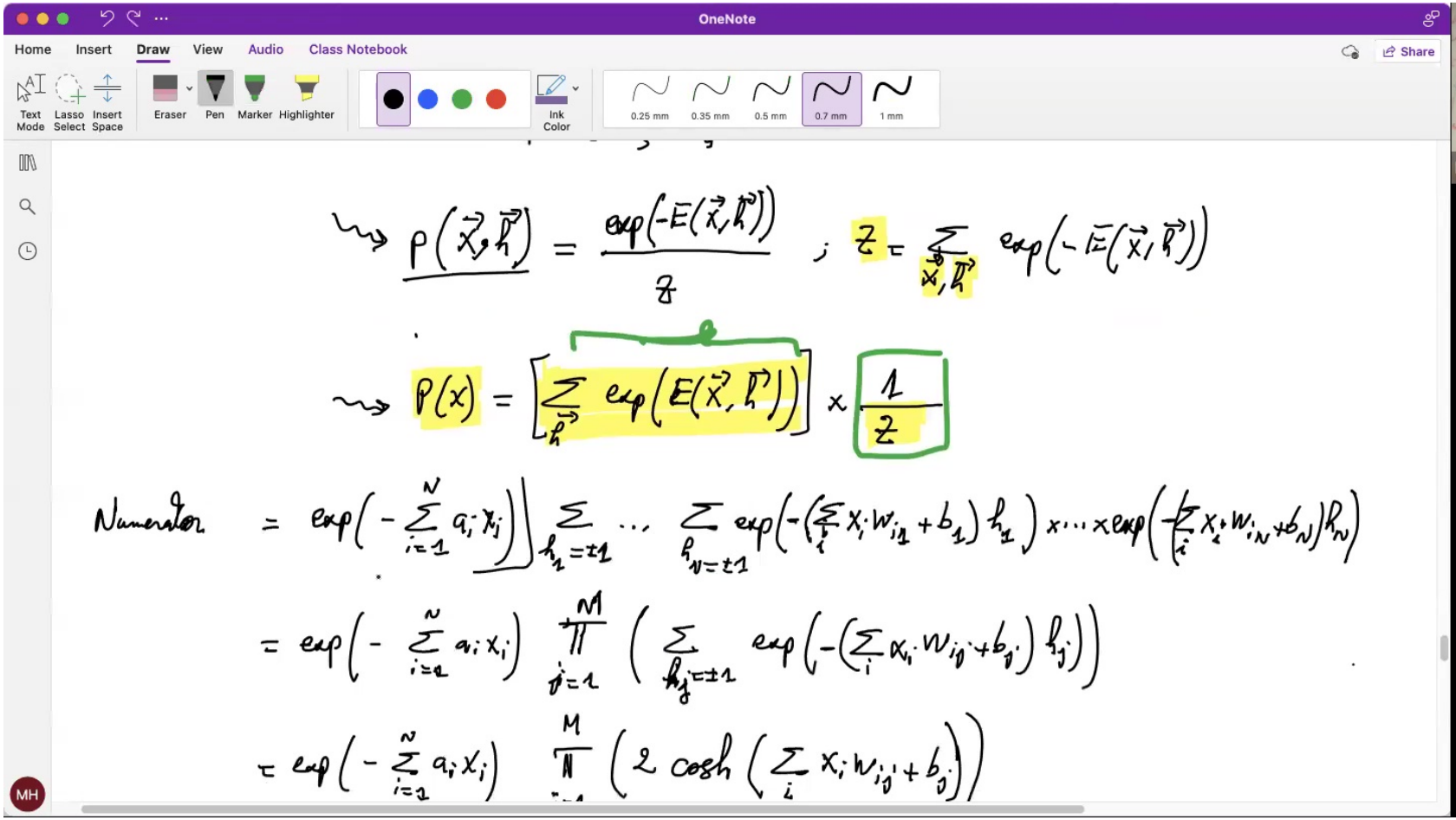
$\vec{h} = (\ ) \updownarrow M$

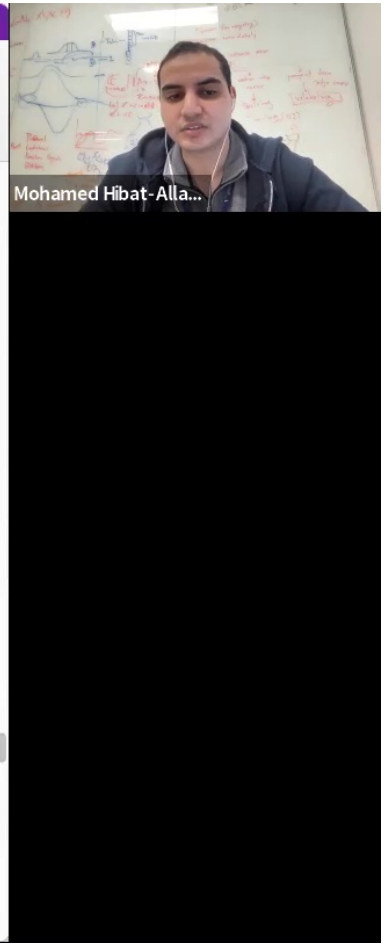


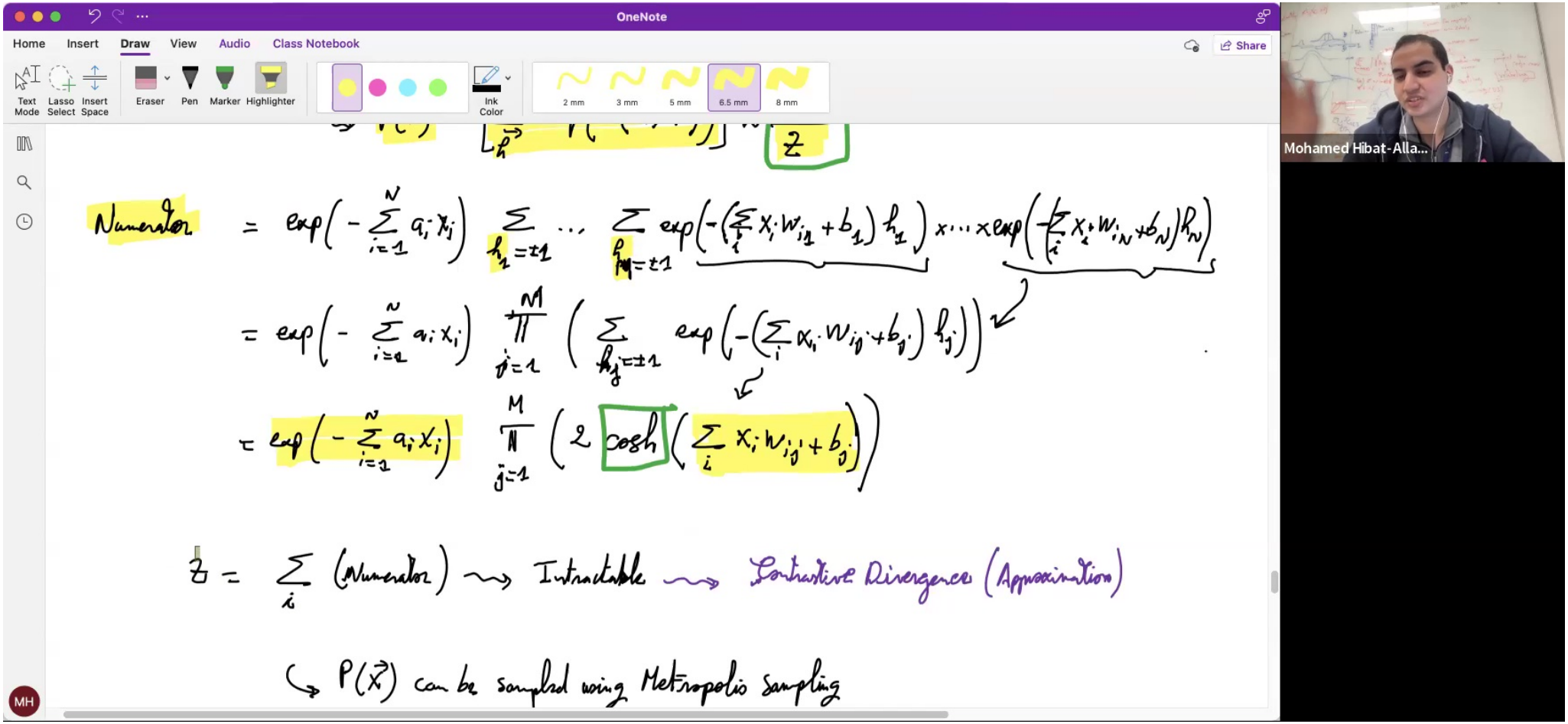
⇝ Correlation between x_1 and x_4 is introduced through h_4.

$$\rightsquigarrow p(\vec{x}, \vec{h}) = \frac{\exp(-E(\vec{x}, \vec{h}))}{Z} \quad ; \quad Z = \sum_{\vec{x}, \vec{h}} \exp(-E(\vec{x}, \vec{h}))$$

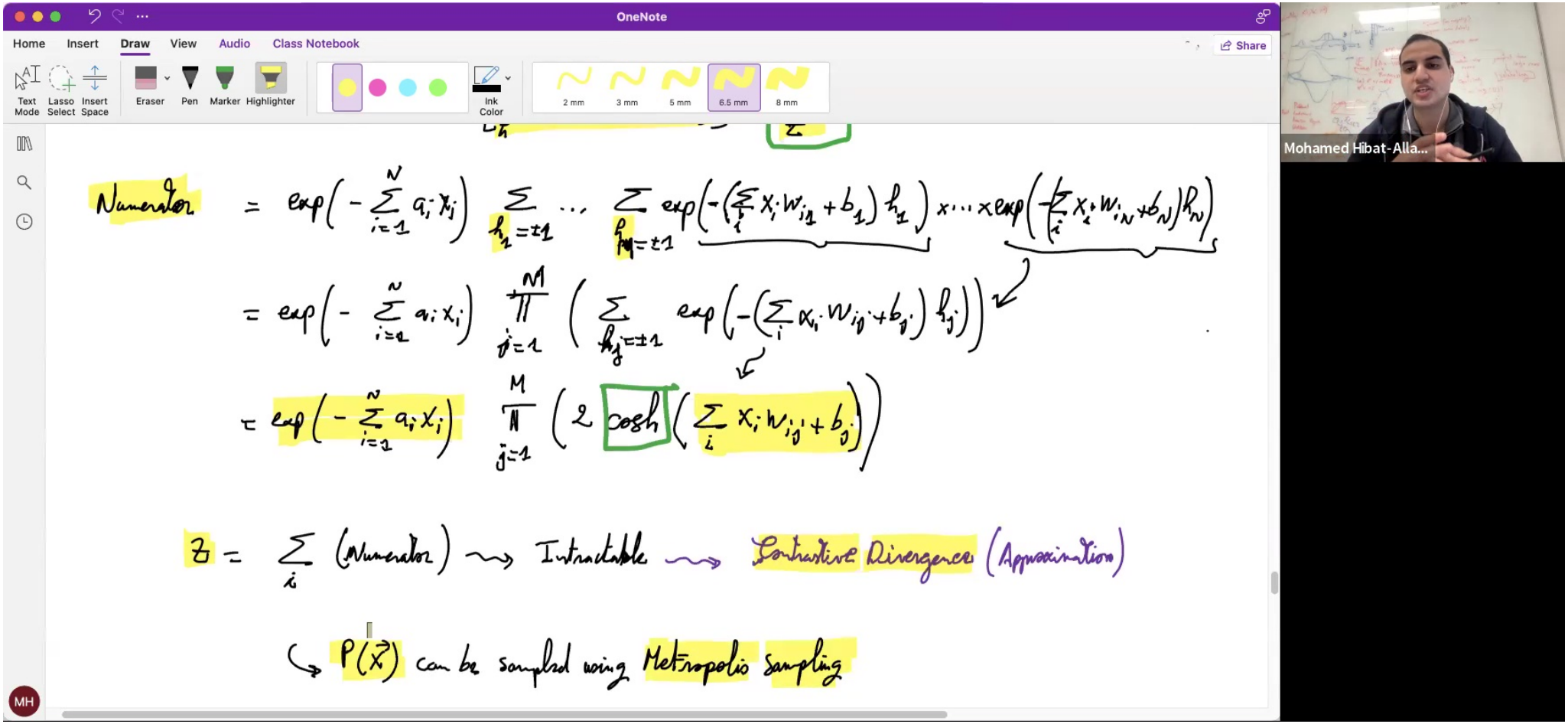
MH


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
0.25 mm
0.35 mm
0.5 mm
0.7 mm
1 mm
$\rightsquigarrow p(\vec{x},\vec{h}) = \frac{\exp(-E(\vec{x},\vec{h}))}{Z}$; $Z = \sum_{\vec{x},\vec{h}} \exp(-E(\vec{x},\vec{h}))$
$\rightsquigarrow p(x) = \left[\sum_{\vec{h}} \exp(E(\vec{x},\vec{h}))\right] \times \frac{1}{Z}$
Numerator $= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \sum_{h_1=\pm 1} \cdots \sum_{h_N=\pm 1} \exp\left(-\left(\sum_i x_i w_{i1} + b_1\right) h_1\right) \times \cdots \times \exp\left(-\left(\sum_i x_i w_{iN} + b_N\right) h_N\right)$
$= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \prod_{j=1}^{M} \left(\sum_{h_j=\pm 1} \exp\left(-\left(\sum_i x_i w_{ij} + b_j\right) h_j\right)\right)$
$= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \prod_{j=1}^{M} \left(2\cosh\left(\sum_i x_i w_{ij} + b_j\right)\right)$
MH


Mohamed Hibat-Alla...


OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
Numerator
Z = Σ (Numerator) ⇝ Intractable ⇝ Contrastive Divergence (Approximation)
↳ P(x⃗) can be sampled using Metropolis sampling
MH

Mohamed Hibat-Alla...

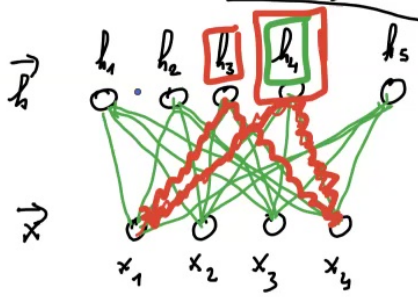

OneNote
Home Insert Draw View Audio Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
2 mm
3 mm
5 mm
6.5 mm
8 mm
Numerator
Contrastive Divergence (Approximation)
Intractable
Metropolis sampling
Mohamed Hibat-Alla...
MH

* Energy-based models ⇝ Restricted Boltzmann Machines (RBM)

$$E(\vec{x}, \vec{h}) = \sum_{i=1}^{N}\sum_{j=1}^{M} x_i w_{ij} h_j - \sum_{i=1}^{N} a_i x_i - \sum_{j=1}^{M} b_j h_j$$

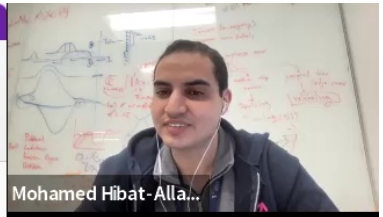
$\vec{x} = ( )^{N}$

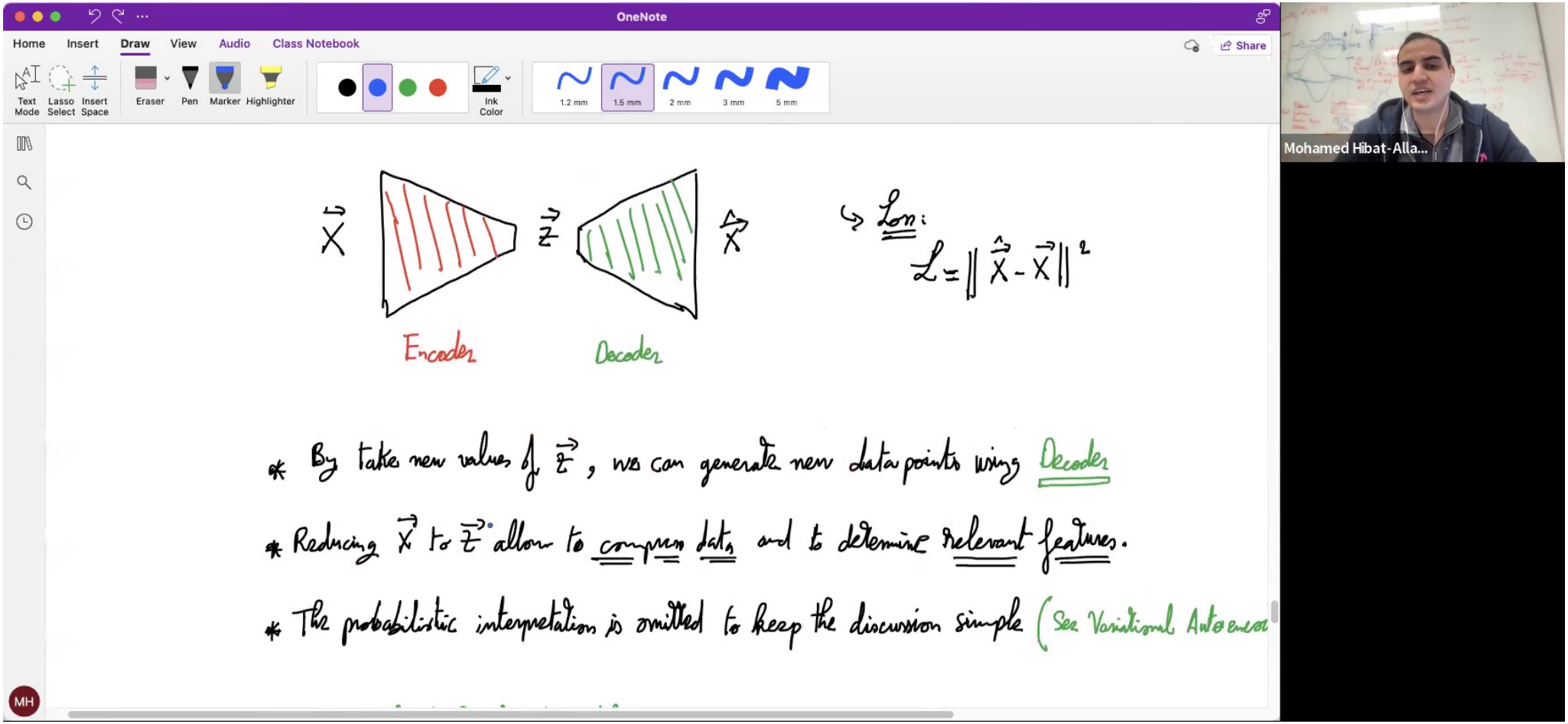
$\vec{h} = ( )^{M}$

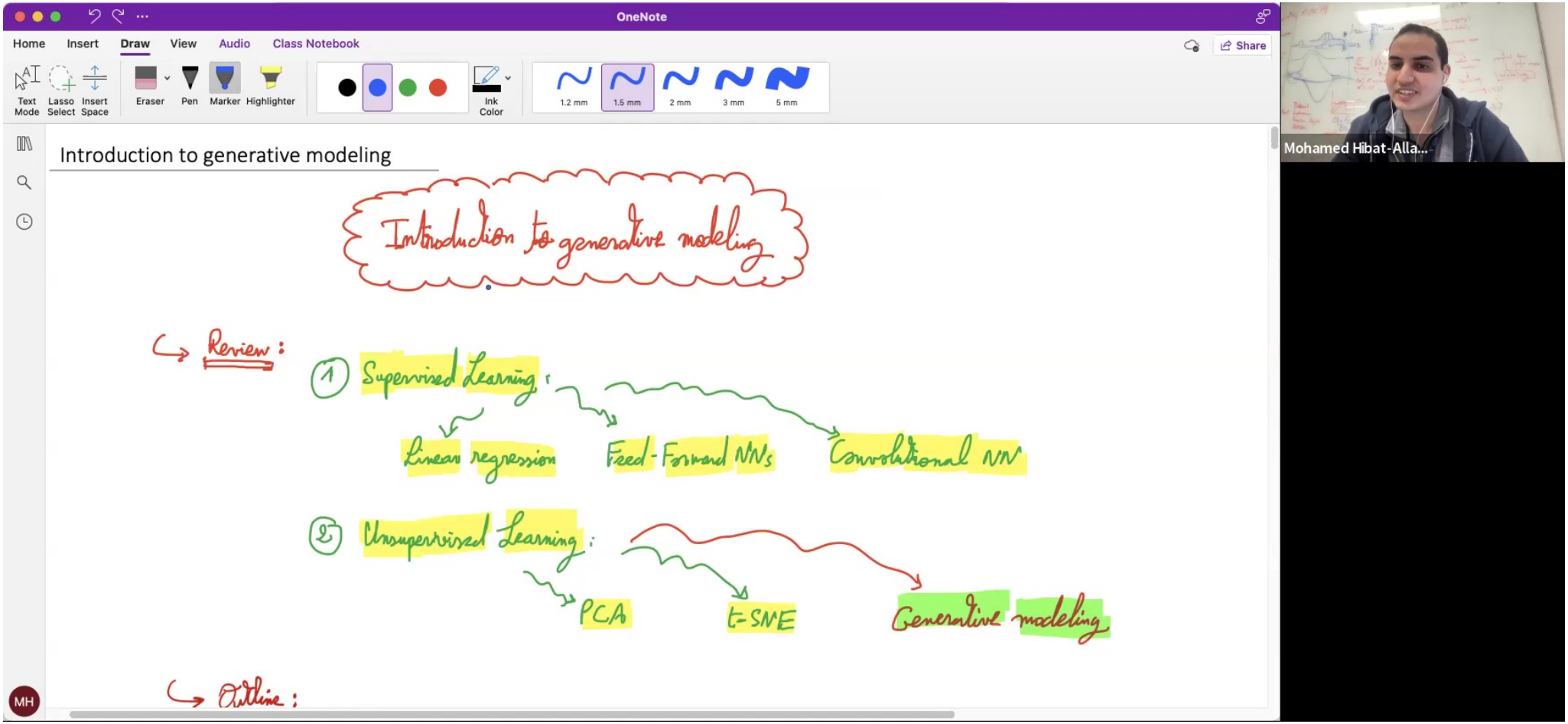


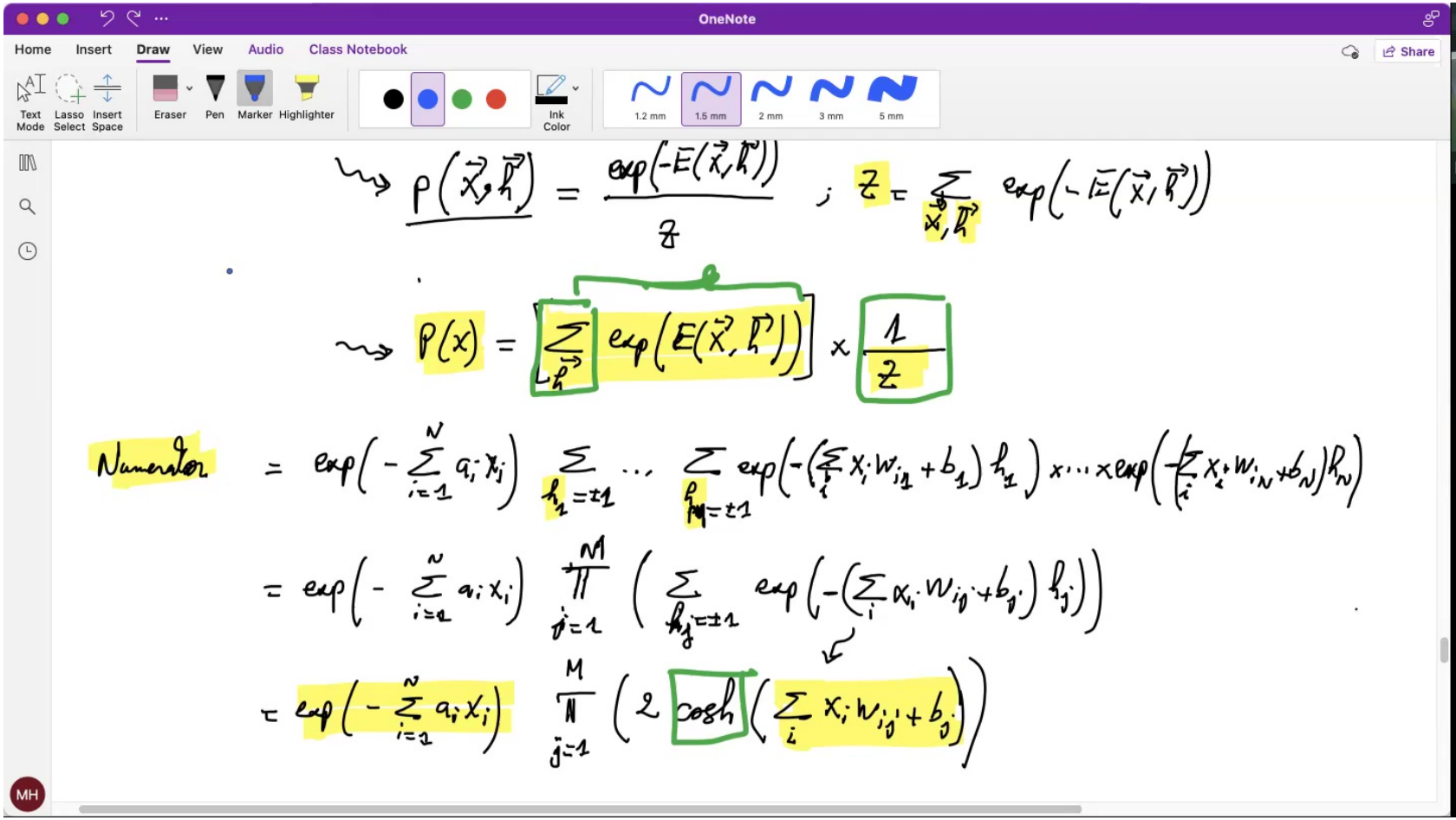
⇝ Correlation between x_1 and x_4 is introduced through h_4.

$$\rightsquigarrow p(\vec{x}, \vec{h}) = \frac{\exp(-E(\vec{x}, \vec{h}))}{Z} \quad ; \quad Z = \sum_{\vec{x}, \vec{h}} \exp(-E(\vec{x}, \vec{h}))$$




OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
Encoder
Decoder
↳ Loss:
$\mathcal{L} = \| \hat{\vec{X}} - \vec{X} \|^2$
* By taking new values of $\vec{z}$, we can generate new data points using Decoder
* Reducing $\vec{X}$ to $\vec{z}$ allow to compress data and to determine relevant features.
* The probabilistic interpretation is omitted to keep the discussion simple (See Variational Autoencoder
Mohamed Hibat-Alla...
MH


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
Introduction to generative modeling
Introduction to generative modeling
Review :
1 Supervised Learning :
Linear regression
Feed-Forward NNs
Convolutional NN
2 Unsupervised Learning :
PCA
t-SNE
Generative modeling
Outline :
MH
Mohamed Hibat-Alla...


OneNote
Home
Insert
Draw
View
Audio
Class Notebook
Share
Text Mode
Lasso Select
Insert Space
Eraser
Pen
Marker
Highlighter
Ink Color
1.2 mm
1.5 mm
2 mm
3 mm
5 mm
$\rightsquigarrow P(\vec{x},\vec{h}) = \frac{\exp(-E(\vec{x},\vec{h}))}{Z}$; $Z = \sum_{\vec{x},\vec{h}} \exp(-E(\vec{x},\vec{h}))$
$\rightsquigarrow P(x) = \left[\sum_{\vec{h}} \exp(E(\vec{x},\vec{h}))\right] \times \frac{1}{Z}$
Numerator
$= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \sum_{h_1=\pm 1} \cdots \sum_{h_M=\pm 1} \exp\left(-\left(\sum_i x_i w_{i1} + b_1\right) h_1\right) \times \cdots \times \exp\left(-\left(\sum_i x_i w_{iN} + b_N\right) h_N\right)$
$= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \prod_{j=1}^{M} \left(\sum_{h_j=\pm 1} \exp\left(-\left(\sum_i x_i w_{ij} + b_j\right) h_j\right)\right)$
$= \exp\left(-\sum_{i=1}^{N} a_i x_i\right) \prod_{j=1}^{M} \left(2 \cosh\left(\sum_i x_i w_{ij} + b_j\right)\right)$
MH


Perimeter-A