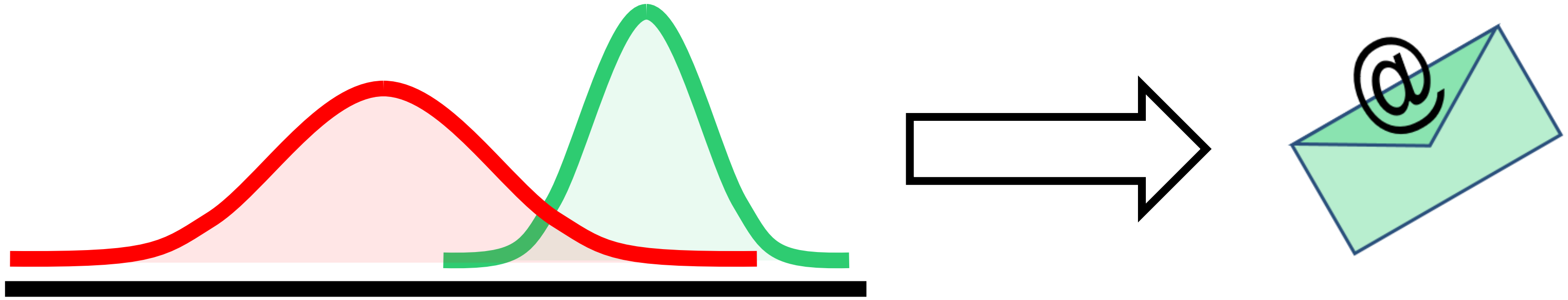
Data Modelling Methods-I

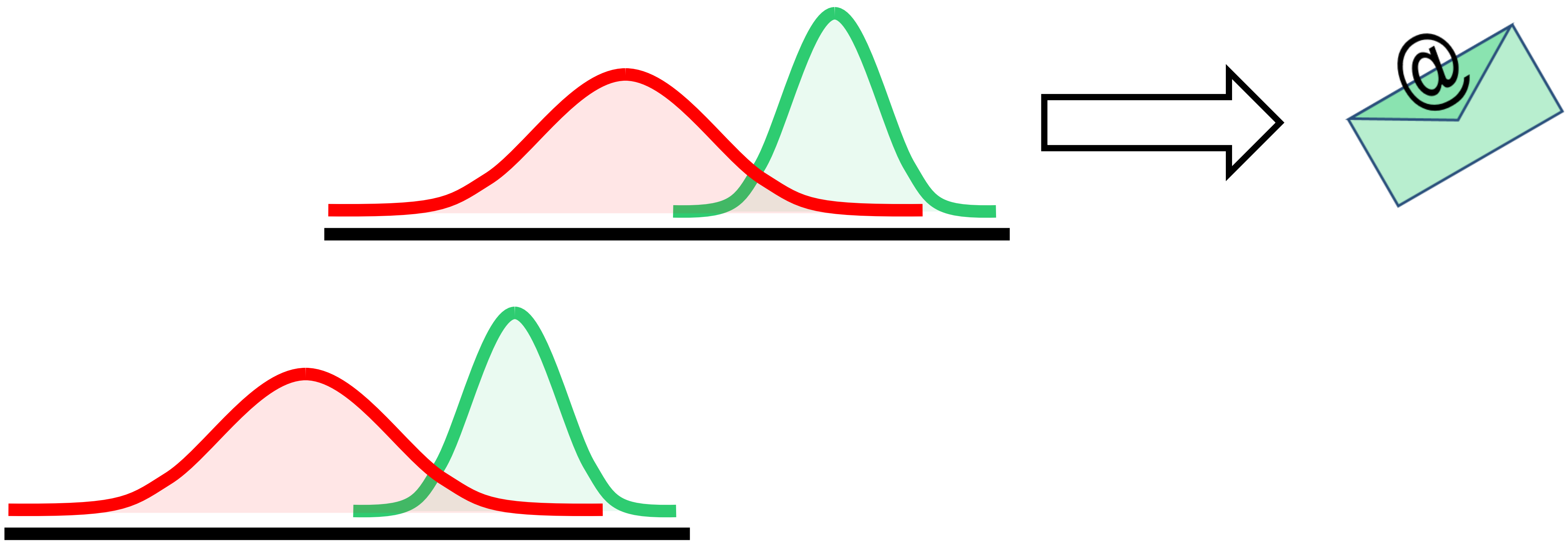
CS771: Introduction to Machine Learning

Purushottam Kar

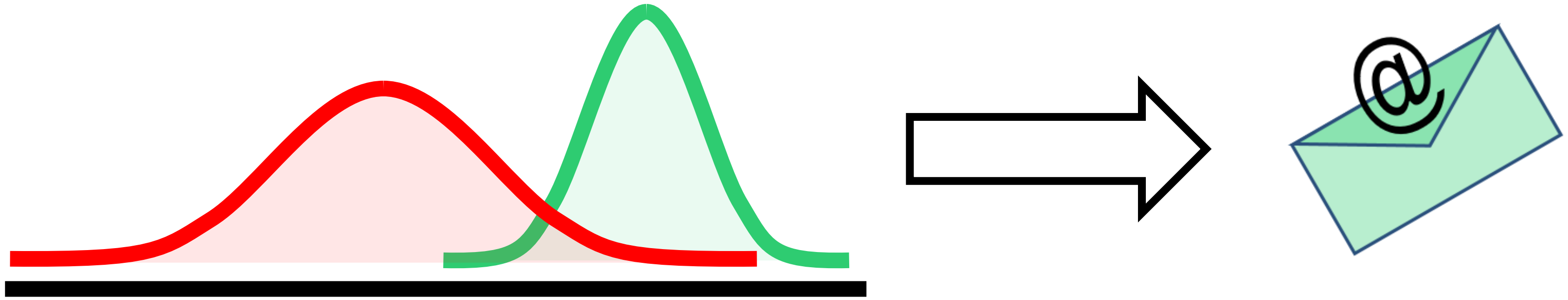
The generative story for labelled data

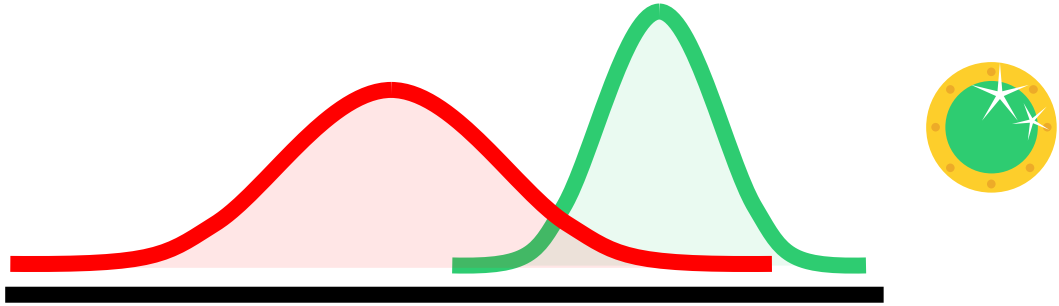


The generative story for labelled data

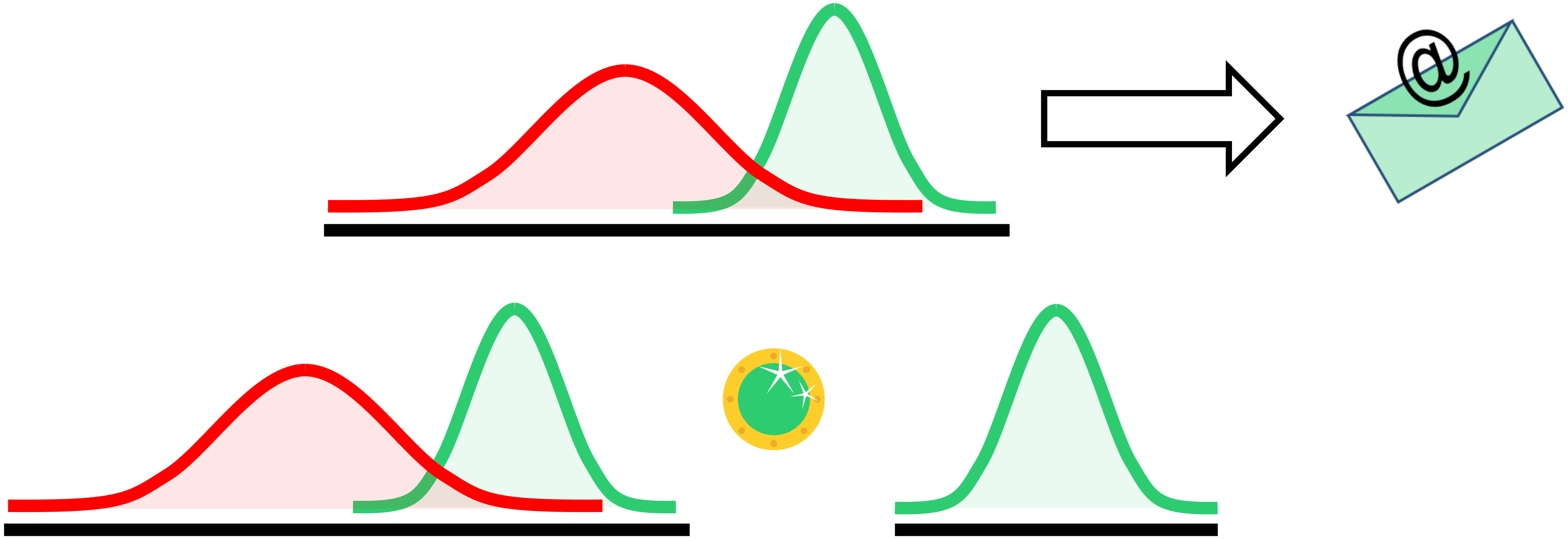


The generative story for labelled data

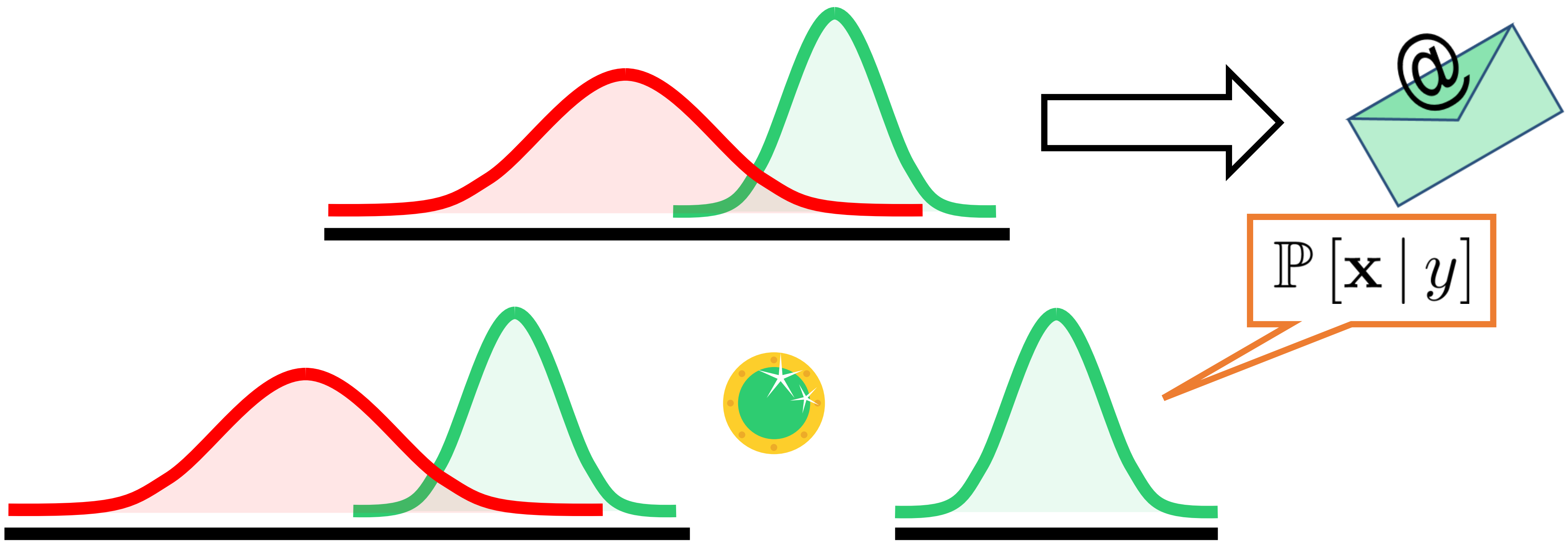




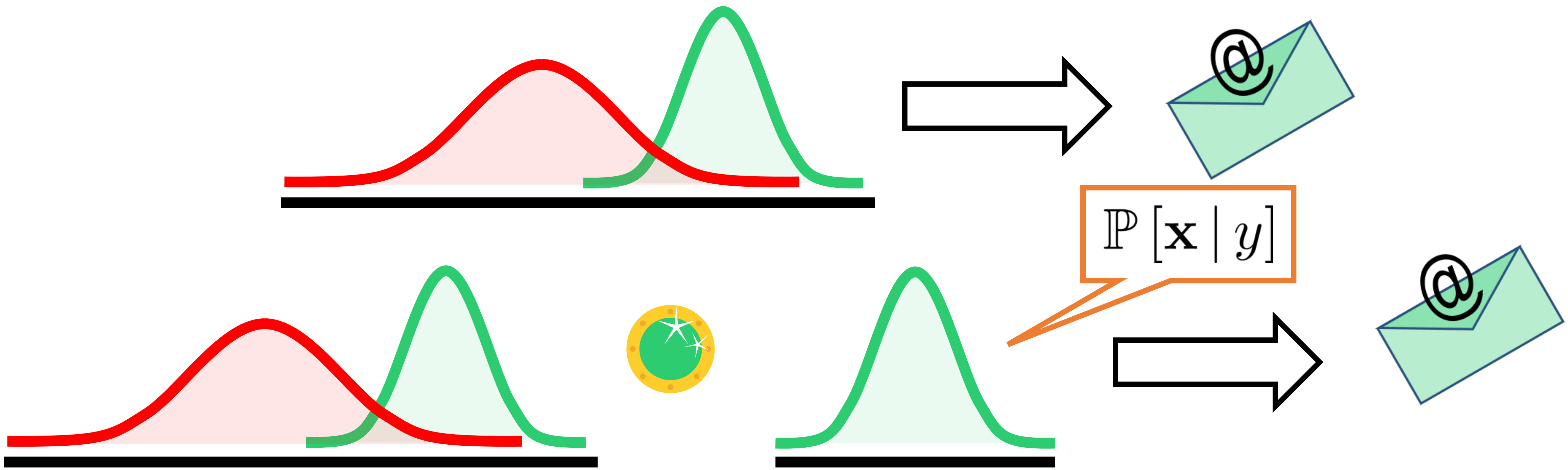
The generative story for labelled data



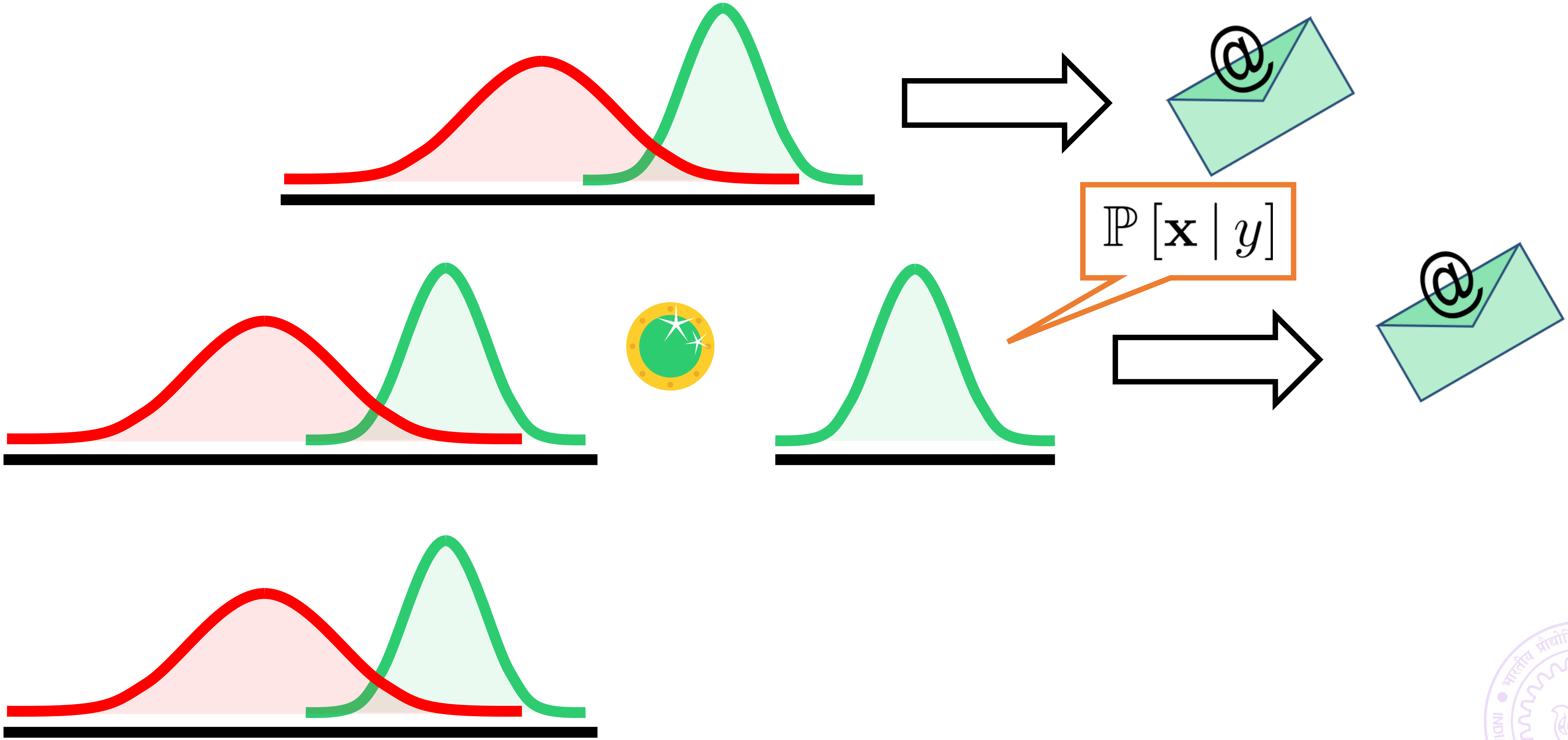
The generative story for labelled data



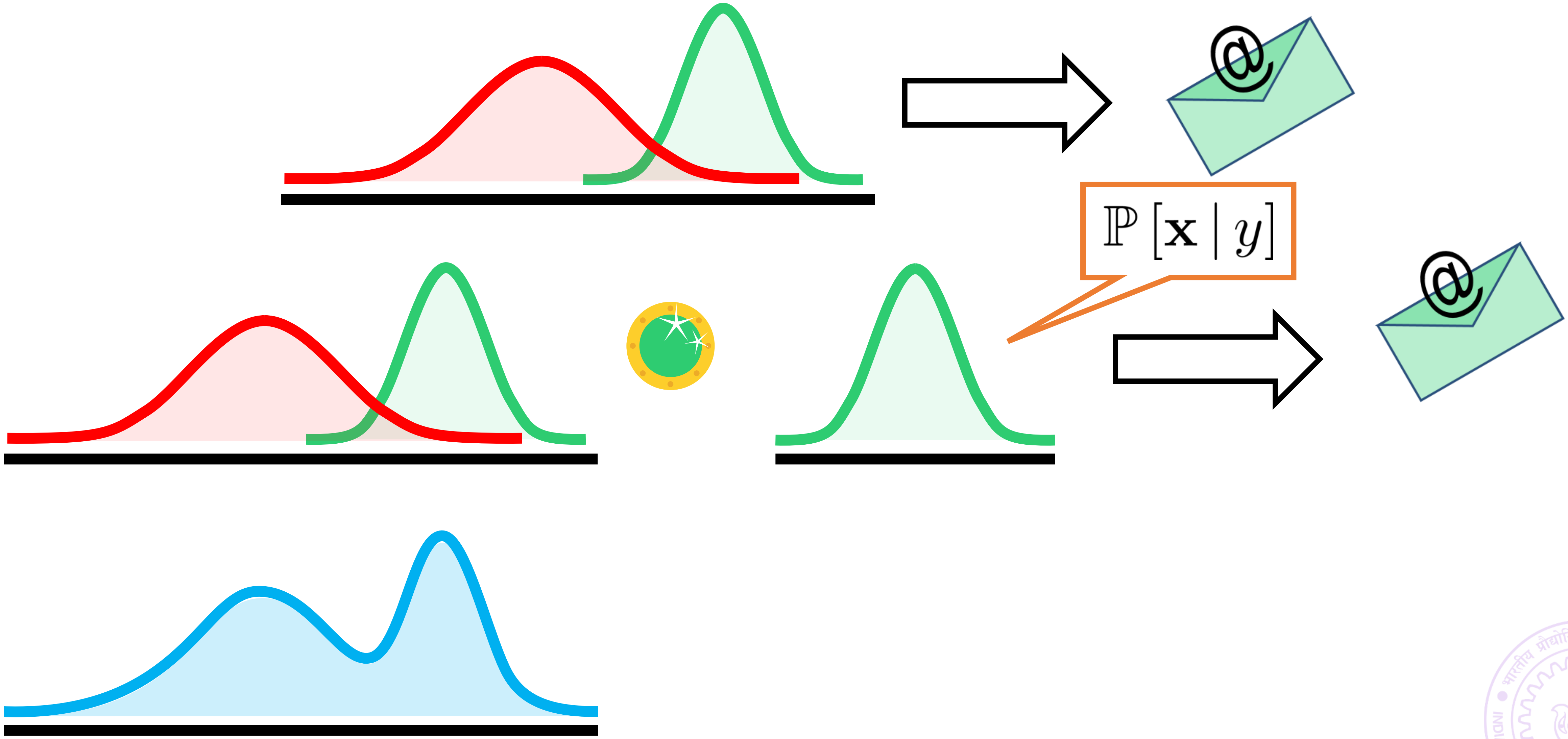
The generative story for labelled data



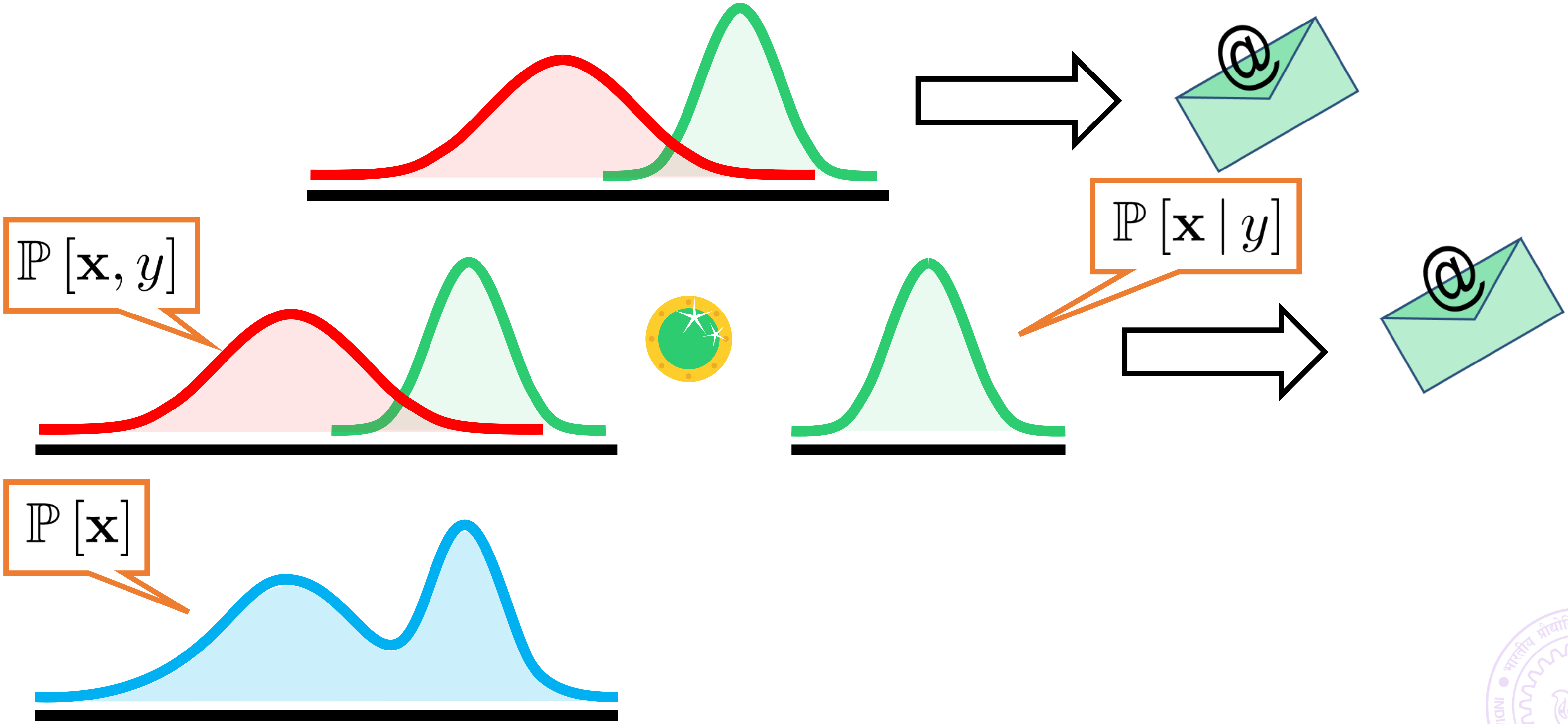
The generative story for labelled data



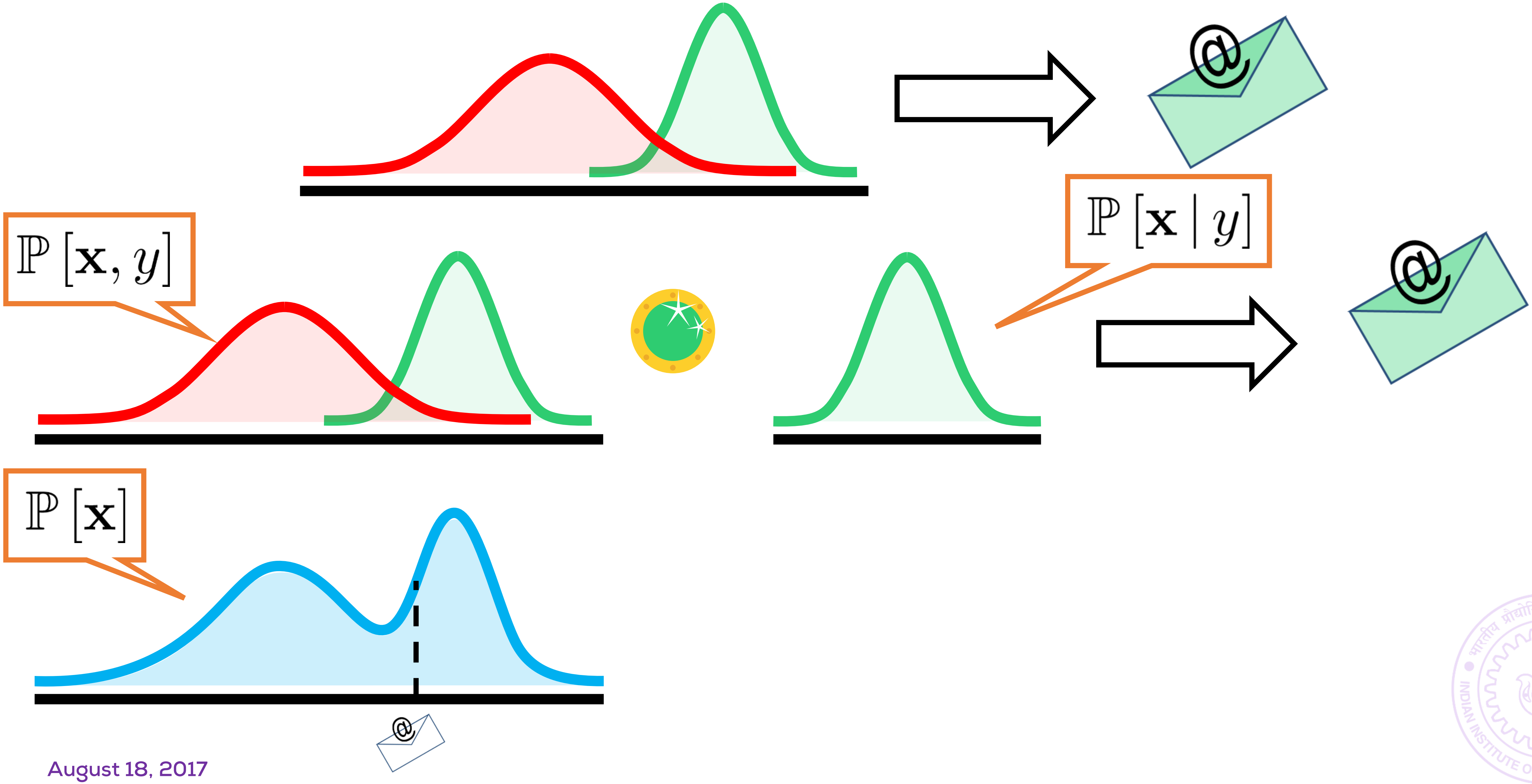
The generative story for labelled data



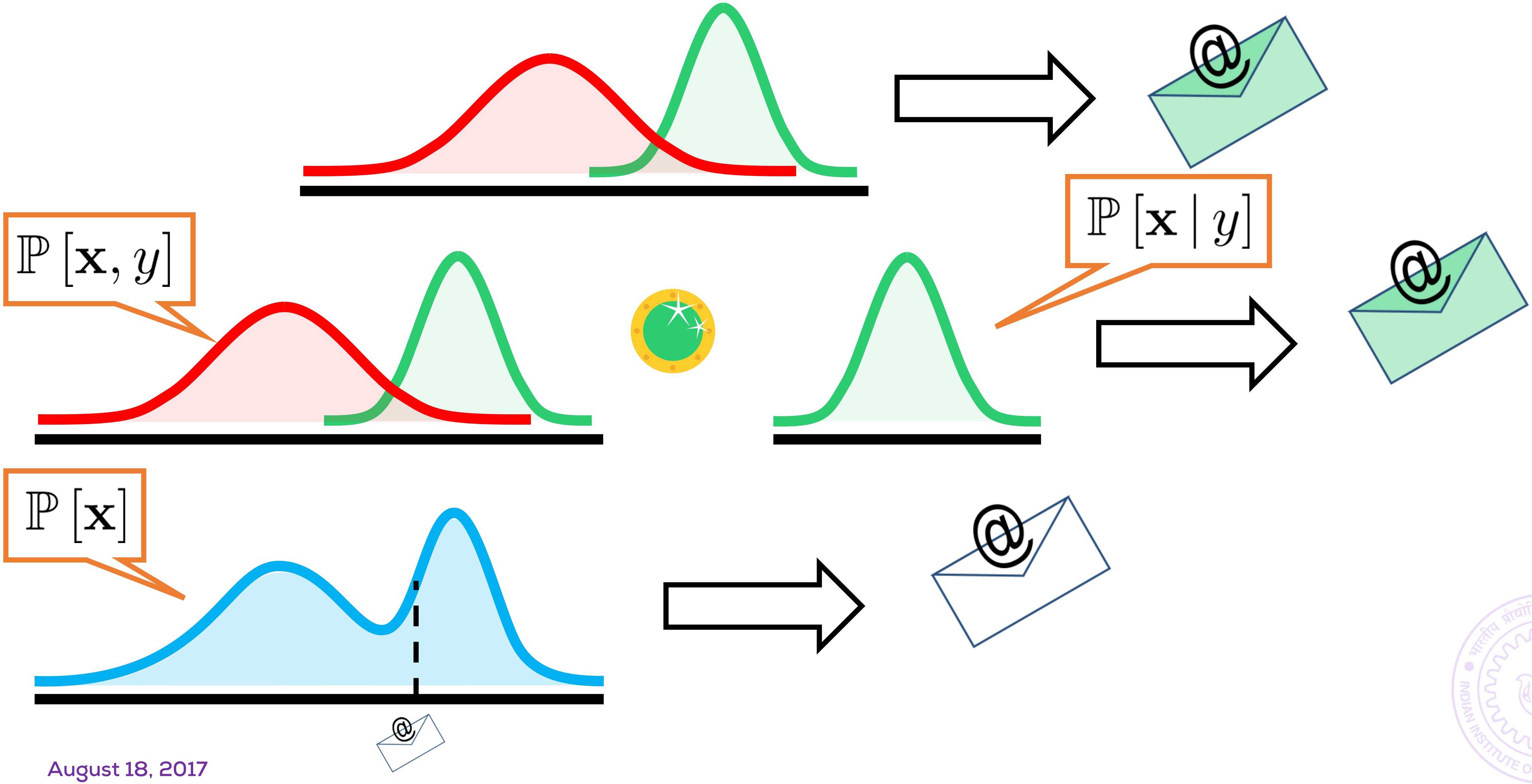
The generative story for labelled data



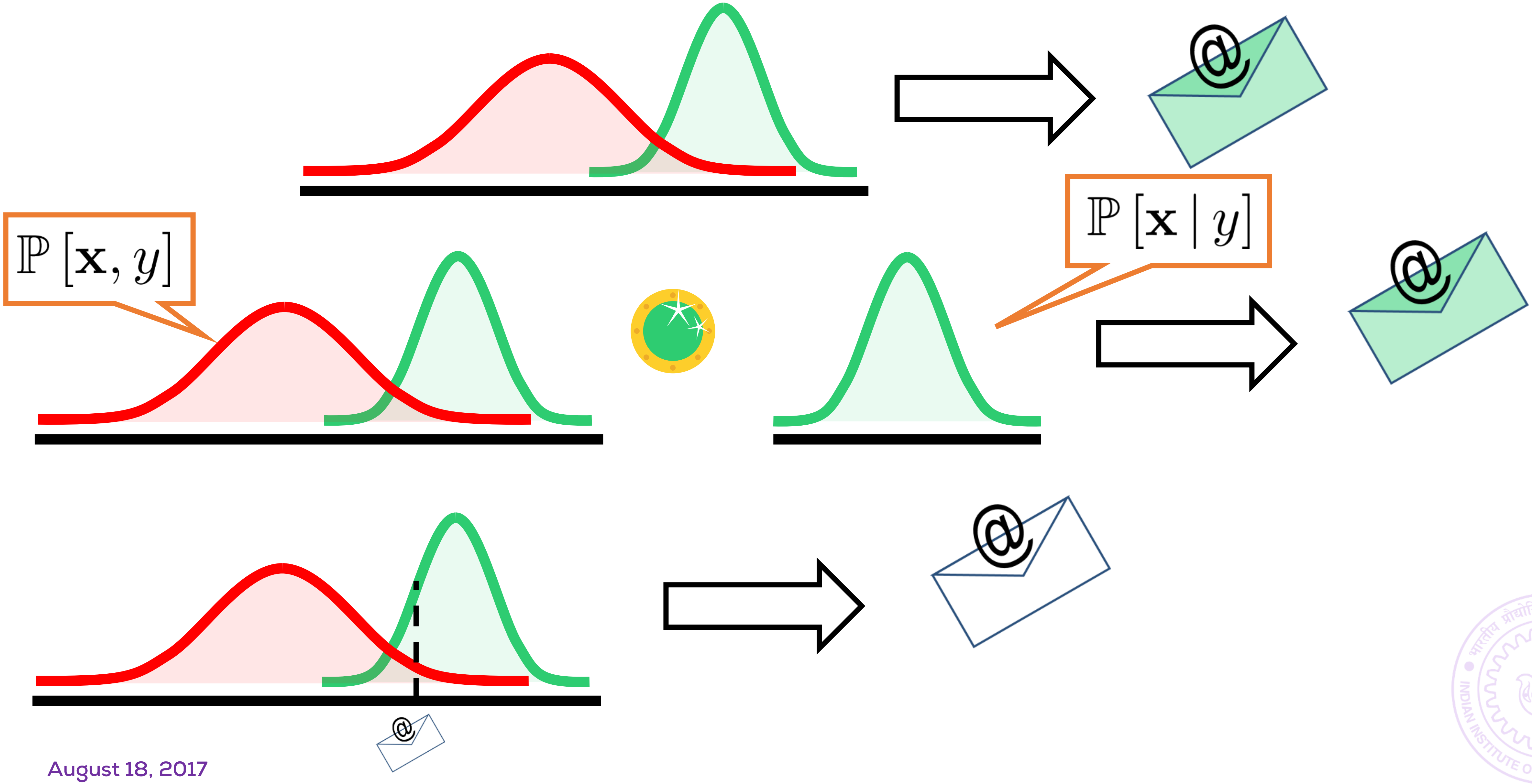
The generative story for labelled data



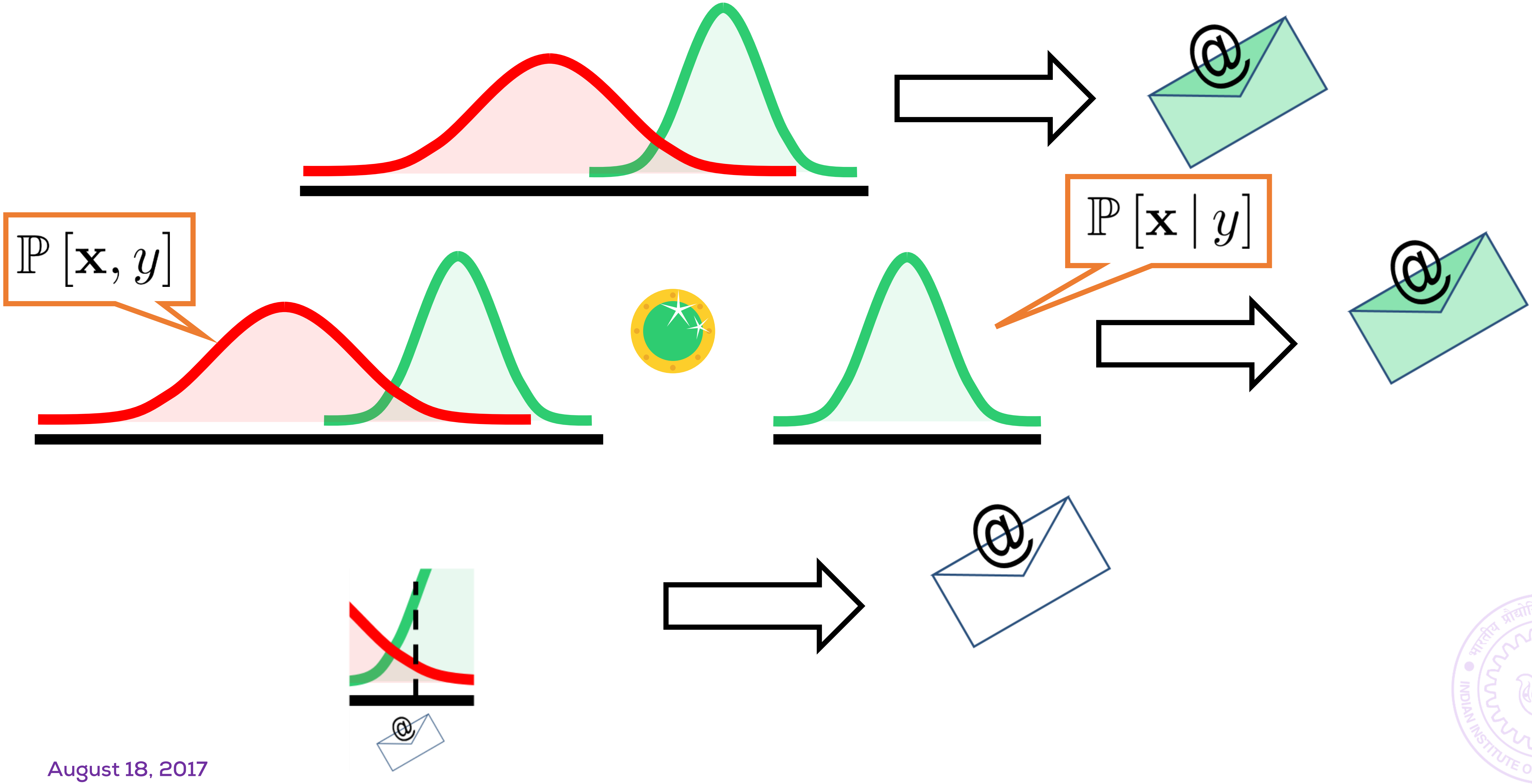
The generative story for labelled data



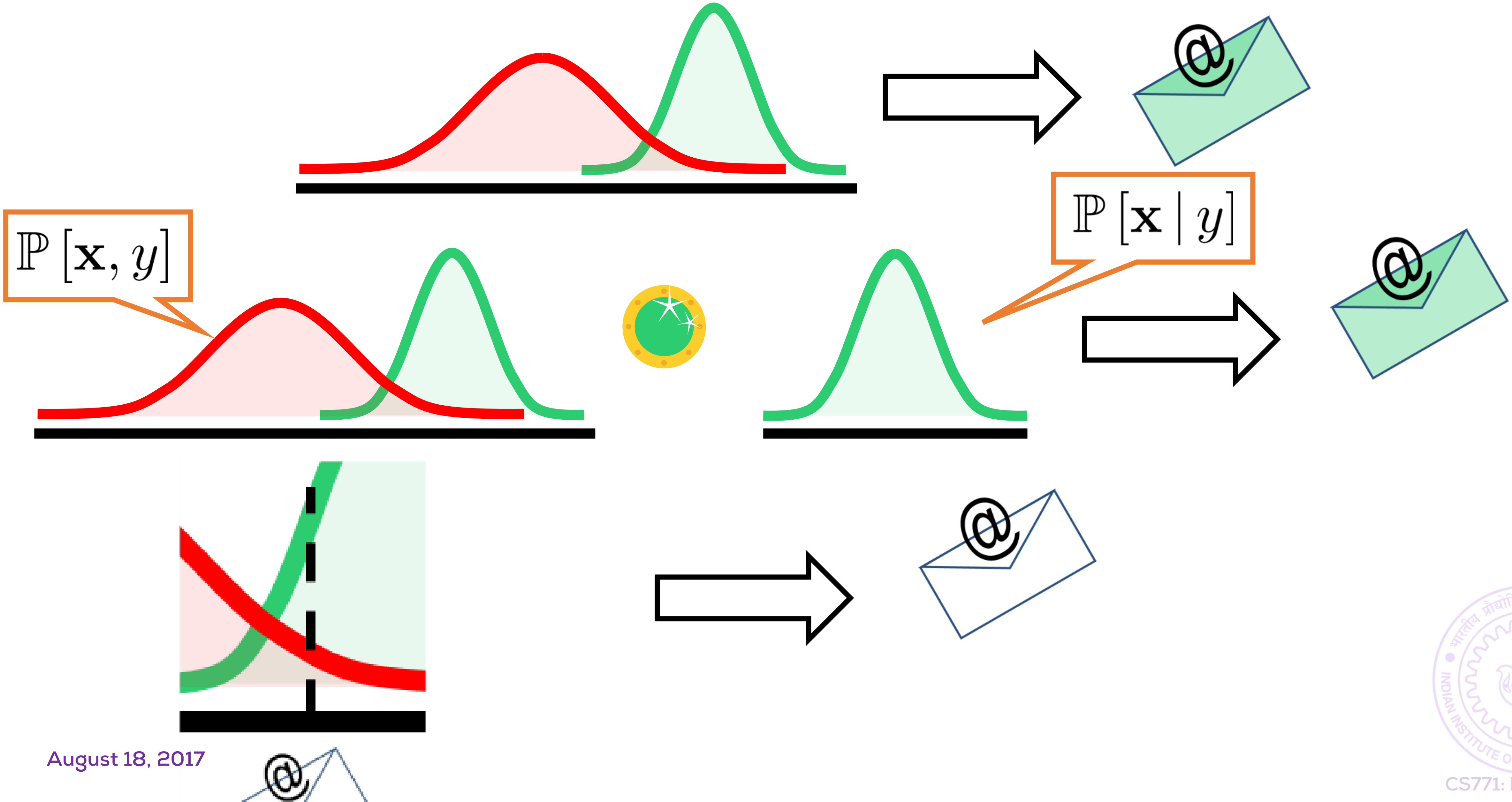
The generative story for labelled data



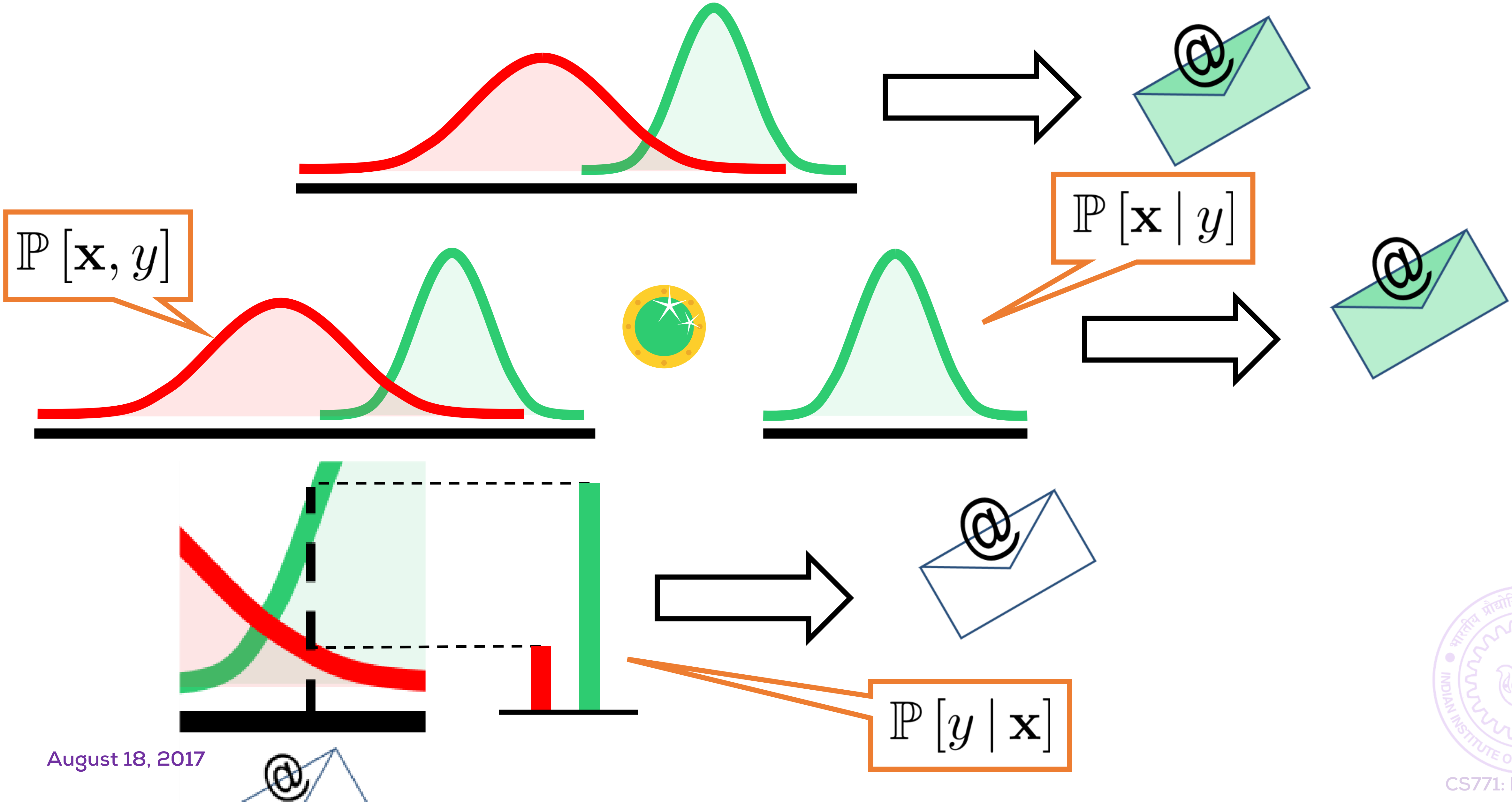
The generative story for labelled data



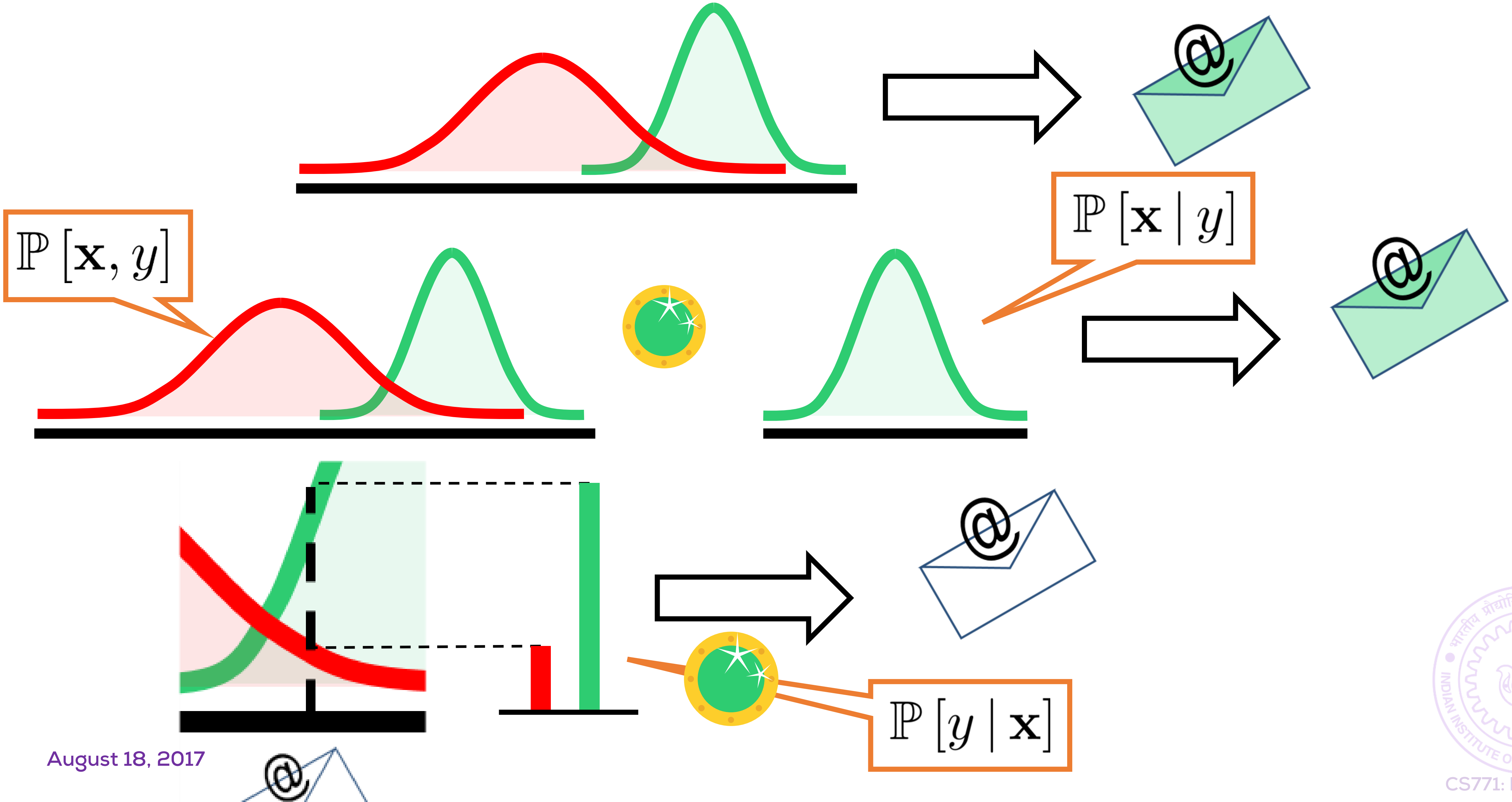
The generative story for labelled data



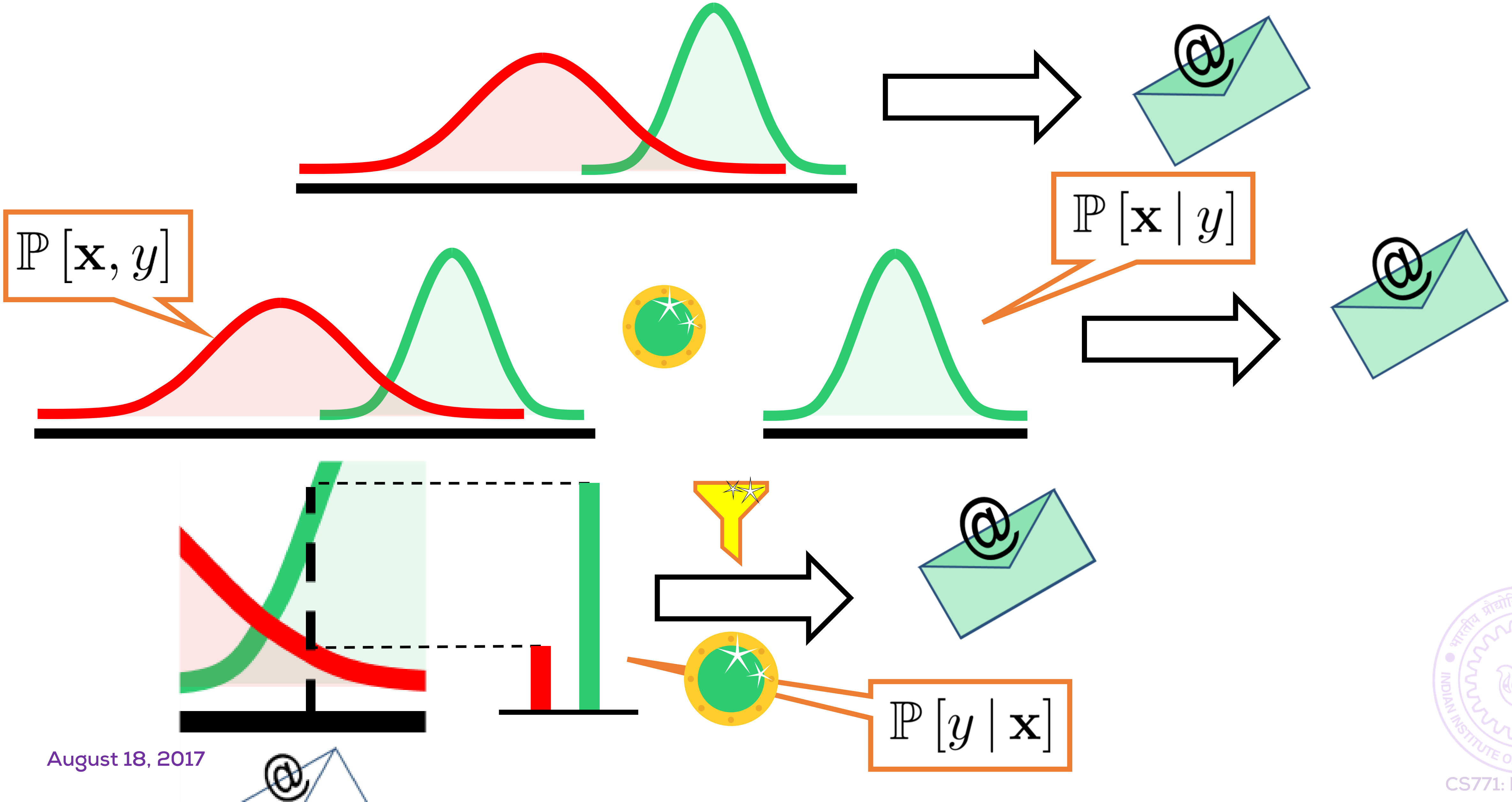
The generative story for labelled data



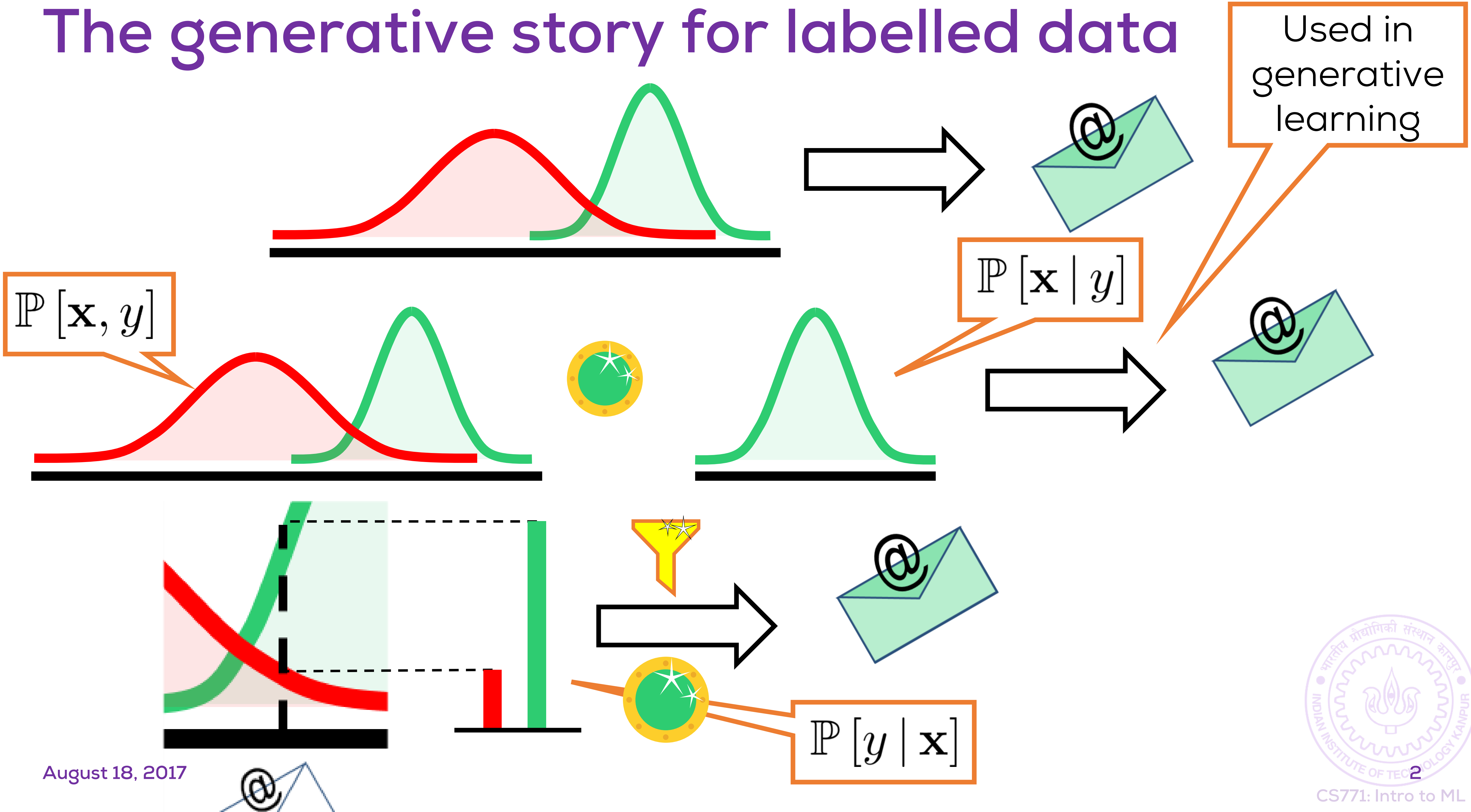
The generative story for labelled data



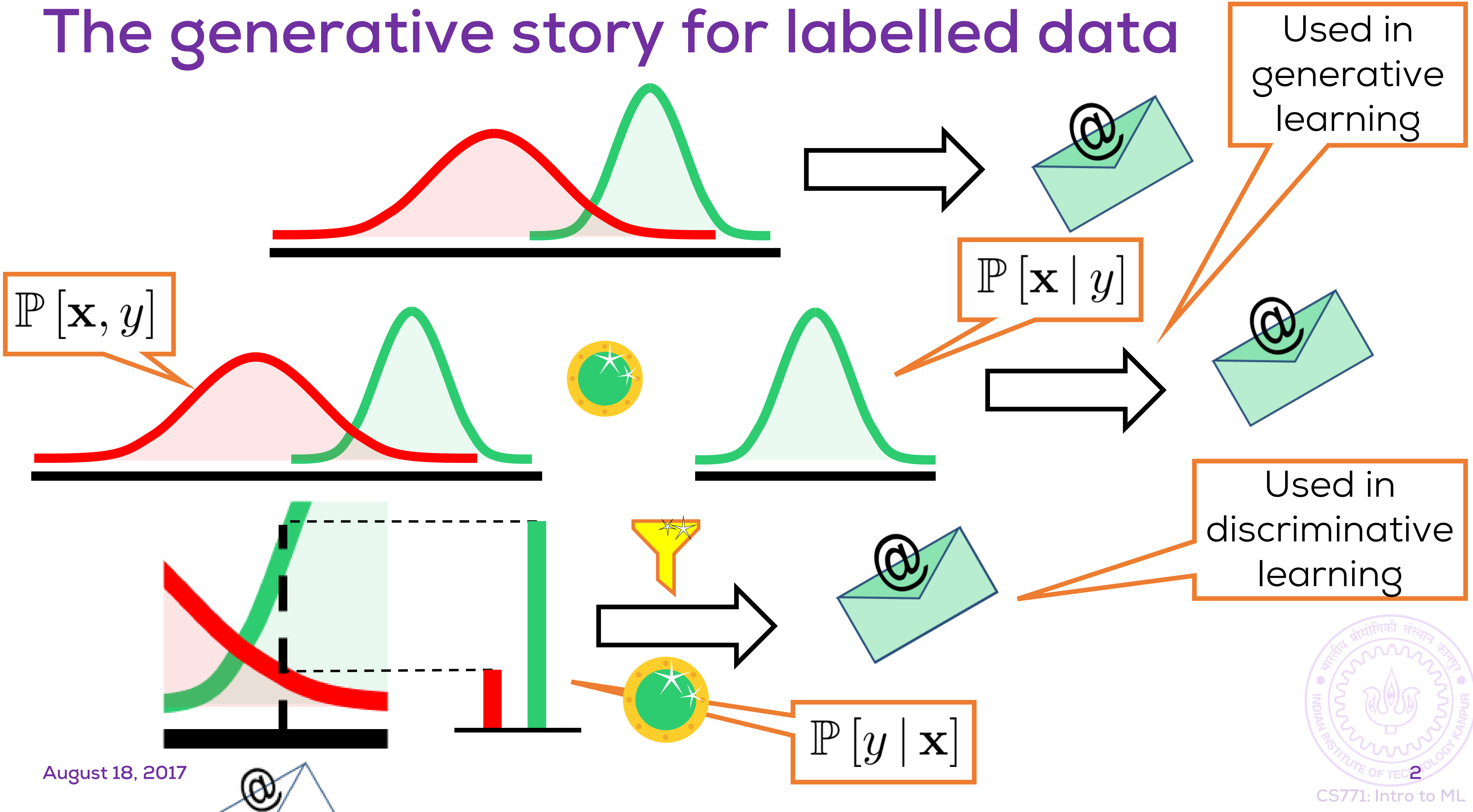
The generative story for labelled data



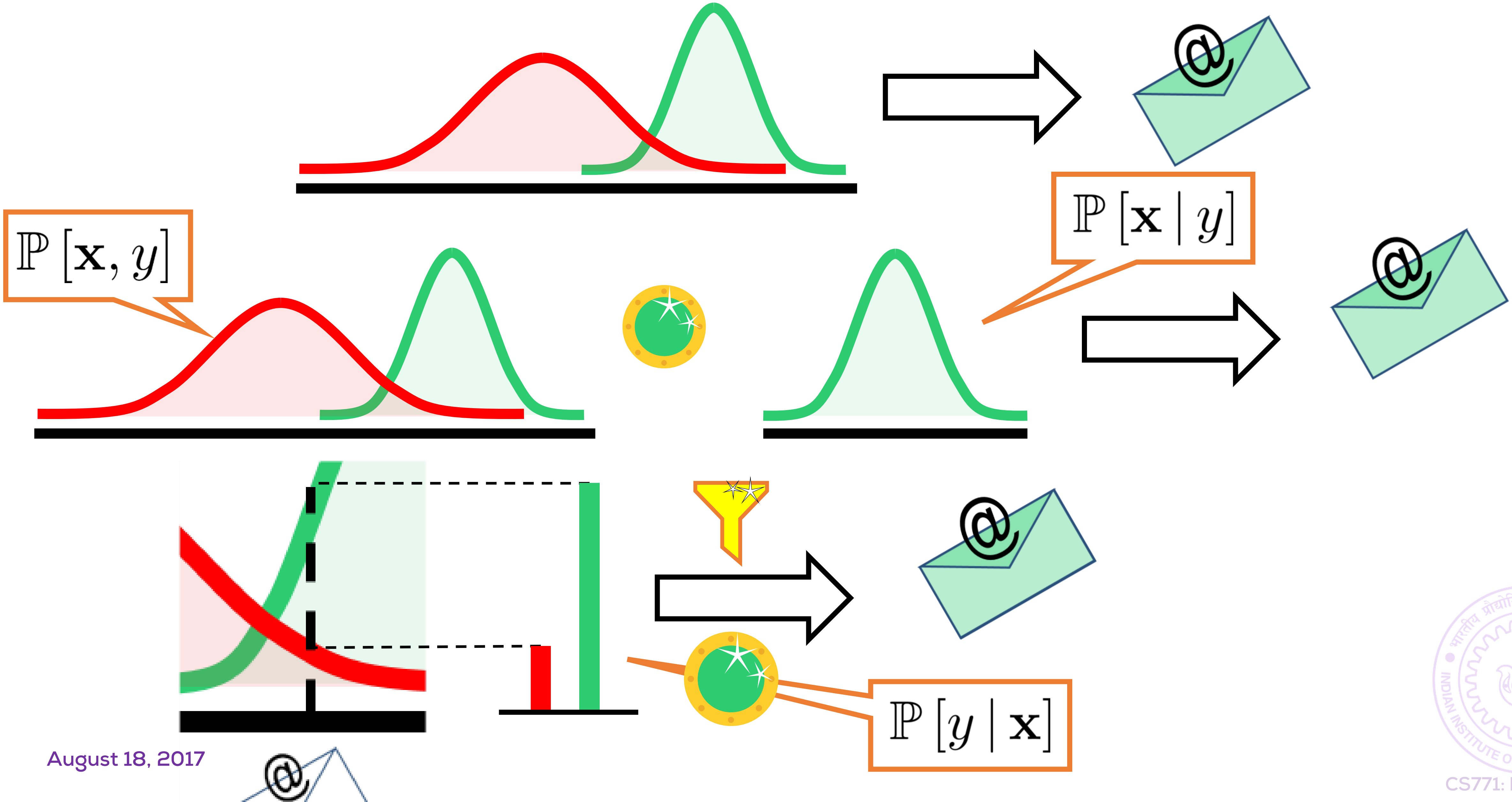
The generative story for labelled data



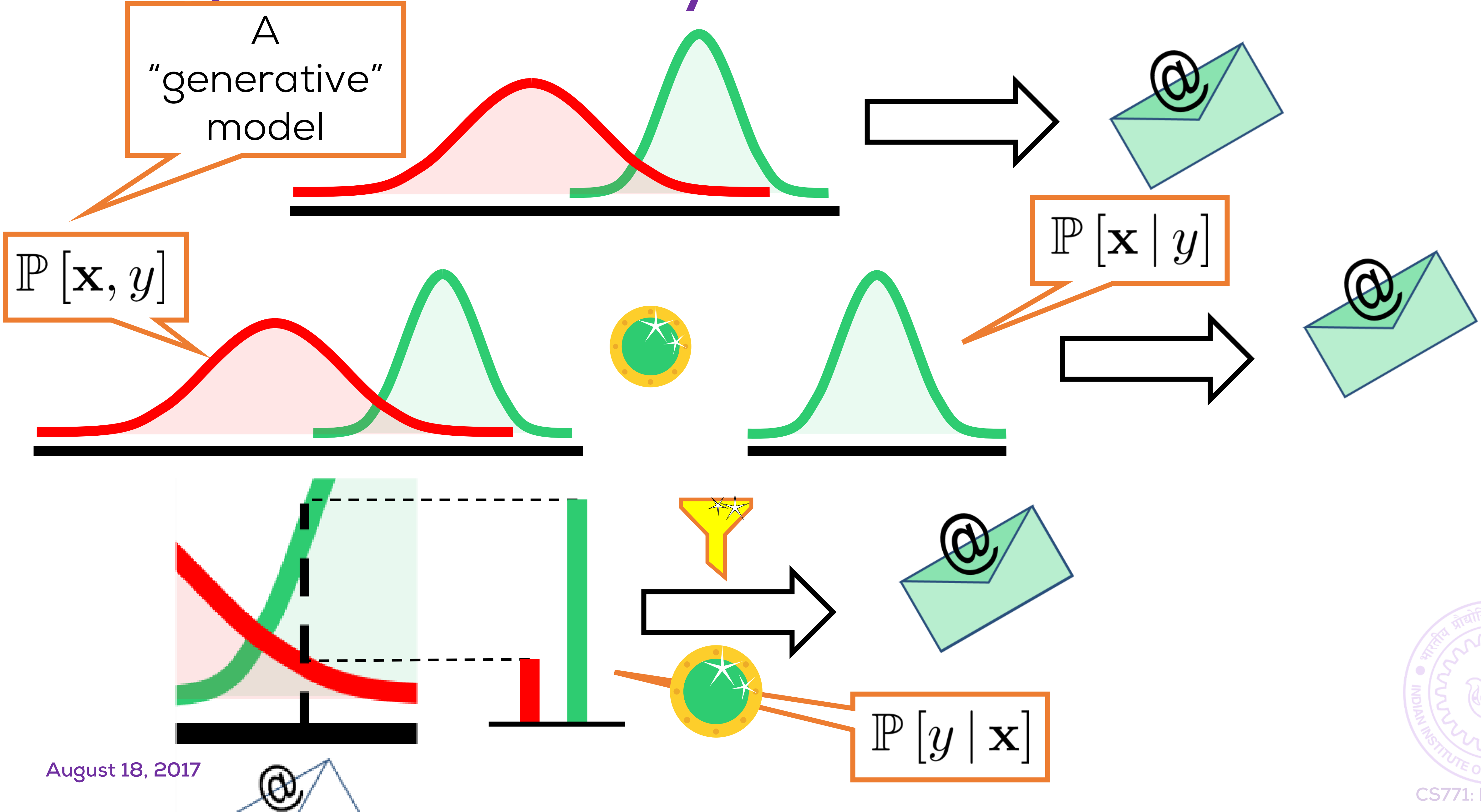
The generative story for labelled data



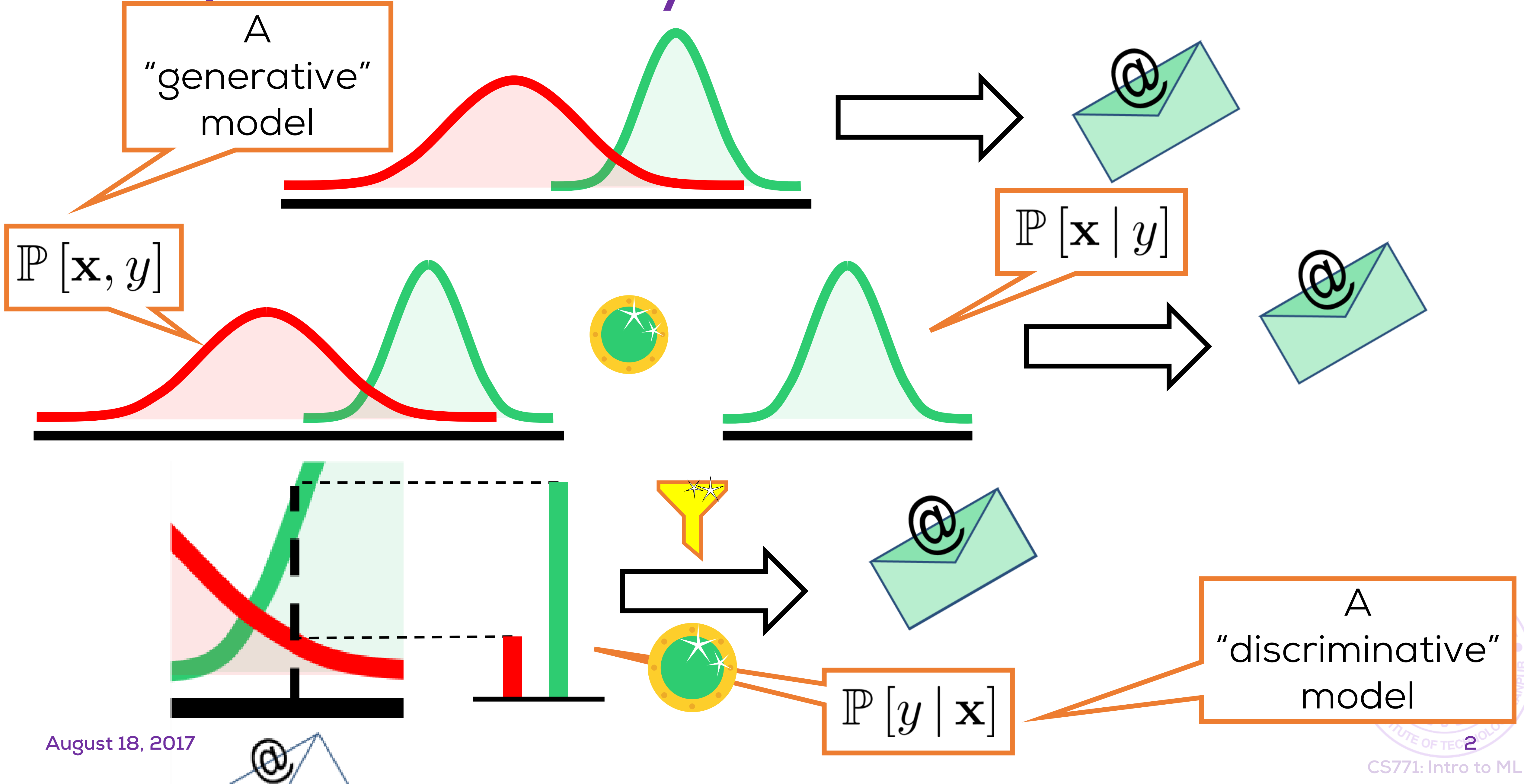
The generative story for labelled data



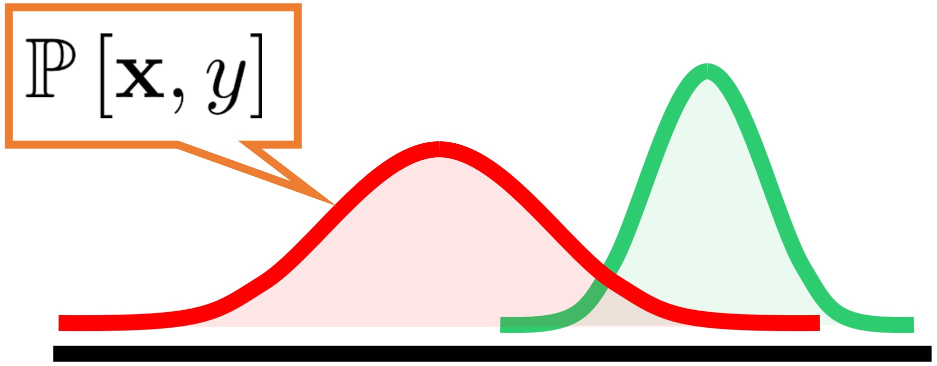
The generative story for labelled data



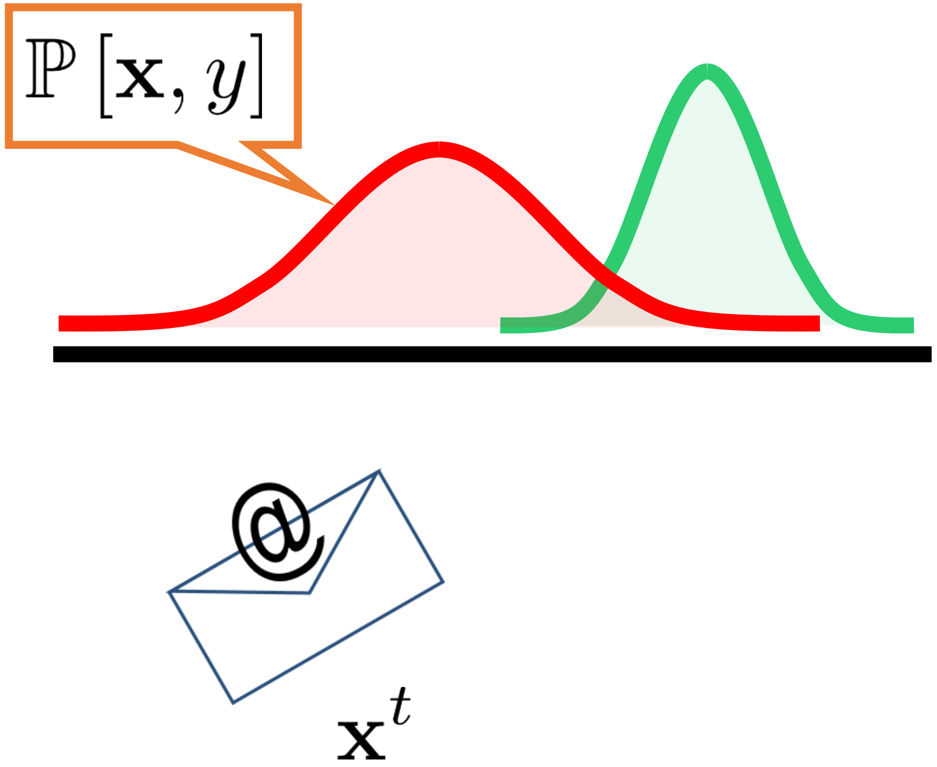
The generative story for labelled data



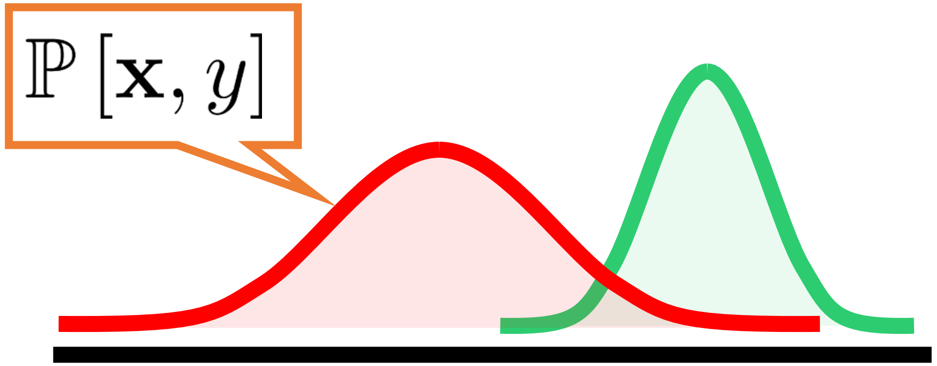
The generative story for labelled data

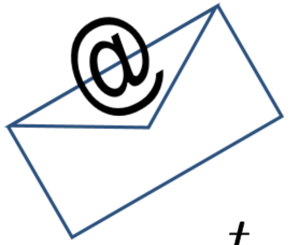


The generative story for labelled data



The generative story for labelled data

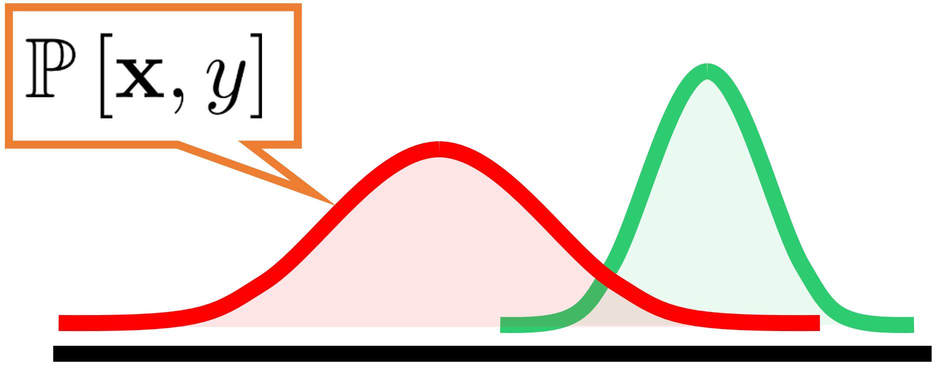


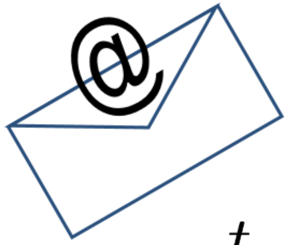


$\mathbf{x}^t$

$$\mathbb{P}\left[y \mid \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

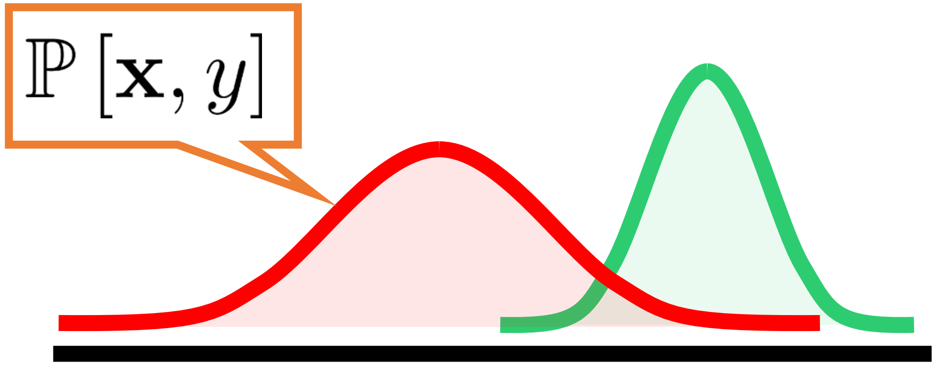
The generative story for labelled data

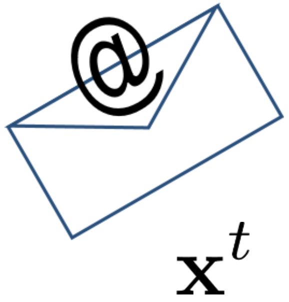




$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]} \qquad \mathbb{P}\left[{\color{red}\bullet} \,|\, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{green}\bullet} \,|\, \mathbf{x}^t\right] \qquad \hat{y}^t = {\color{red}\bullet}$$

The generative story for labelled data





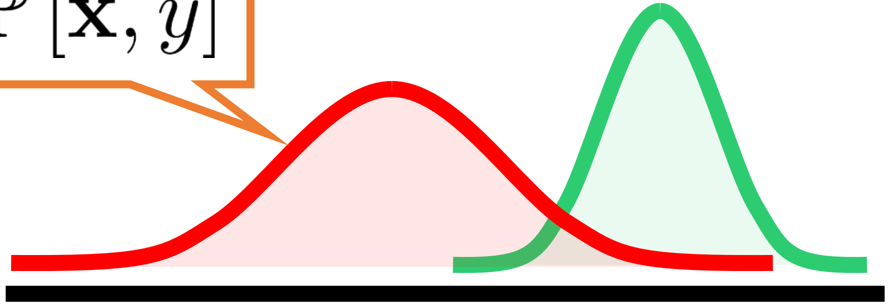
$$\mathbb{P}\left[y \mid \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

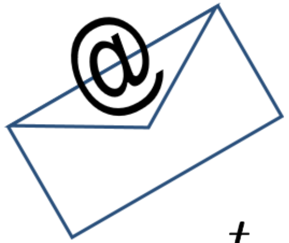
$$\mathbb{P}\left[{\color{red}\bullet} \mid \mathbf{x}^t\right] > \mathbb{P}\left[{\color{green}\bullet} \mid \mathbf{x}^t\right] \quad \hat{y}^t = {\color{red}\bullet}$$

$$\mathbb{P}\left[{\color{green}\bullet} \mid \mathbf{x}^t\right] > \mathbb{P}\left[{\color{red}\bullet} \mid \mathbf{x}^t\right] \quad \hat{y}^t = {\color{green}\bullet}$$

The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$





$\mathbf{x}^t$

$$\mathbb{P}\left[y \mid \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

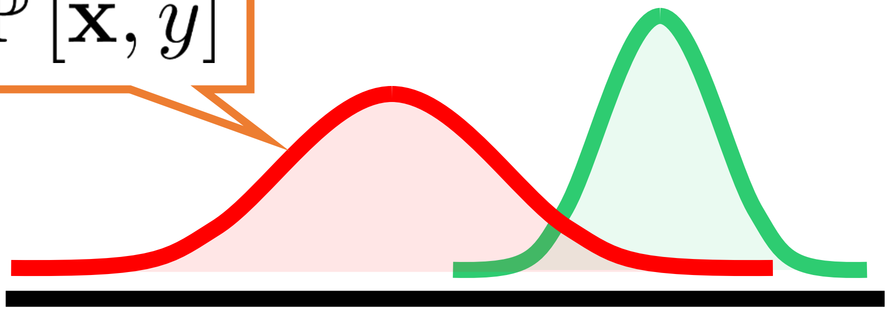
Predict the most likely label

$$\mathbb{P}\left[\textcolor{red}{\bullet} \mid \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{green}{\bullet} \mid \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{red}{\bullet}$$

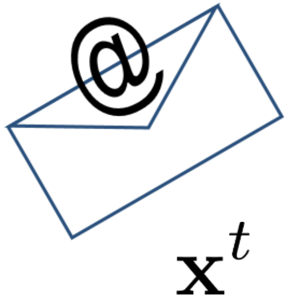
$$\mathbb{P}\left[\textcolor{green}{\bullet} \mid \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{red}{\bullet} \mid \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{green}{\bullet}$$

The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



Predict the most likely label



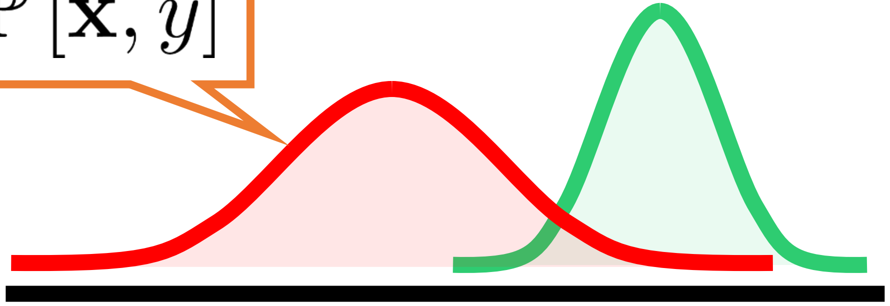
$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{red}{\bullet}$$

$$\mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{green}{\bullet}$$

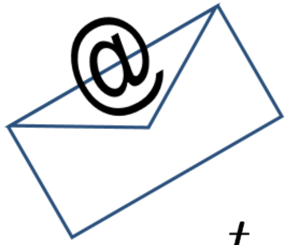
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label



$\mathbf{x}^t$

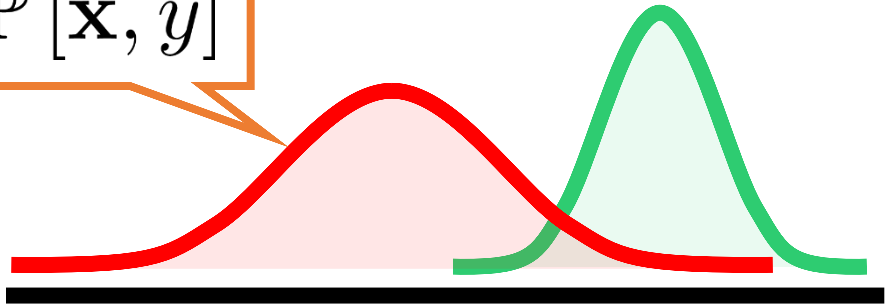
$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{red}\bullet}$$

$$\mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{green}\bullet}$$

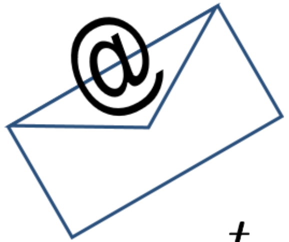
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label



$\mathbf{x}^t$

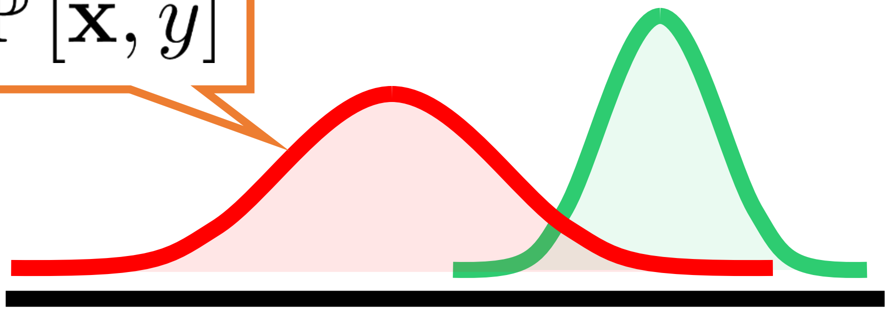
$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{red}{\bullet}$$

$$\mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{green}{\bullet}$$

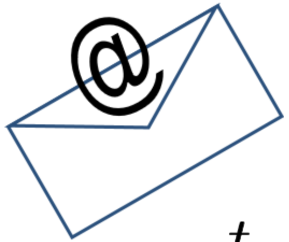
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label

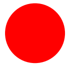


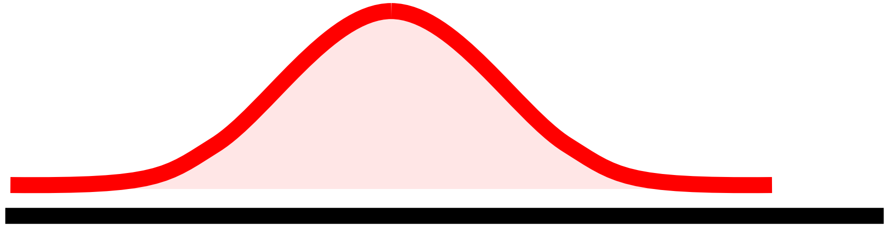
$\mathbf{x}^t$

$$\mathbb{P}\left[y \mid \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{red}\bullet}$$

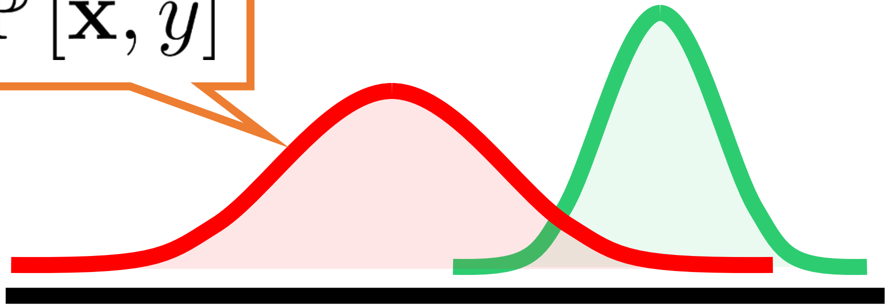
$$\mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{green}\bullet}$$





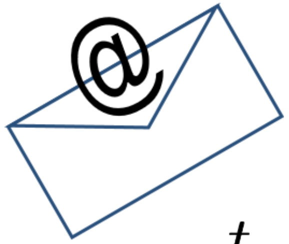
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label

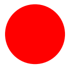


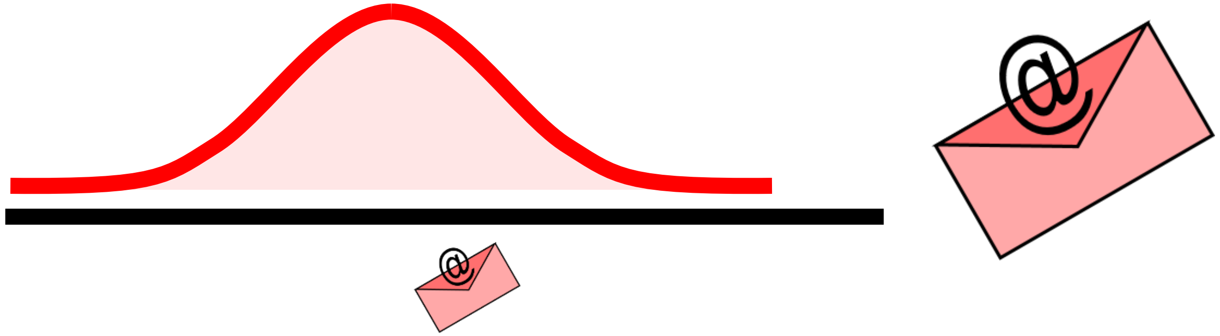
$\mathbf{x}^t$

$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{red}{\bullet}$$

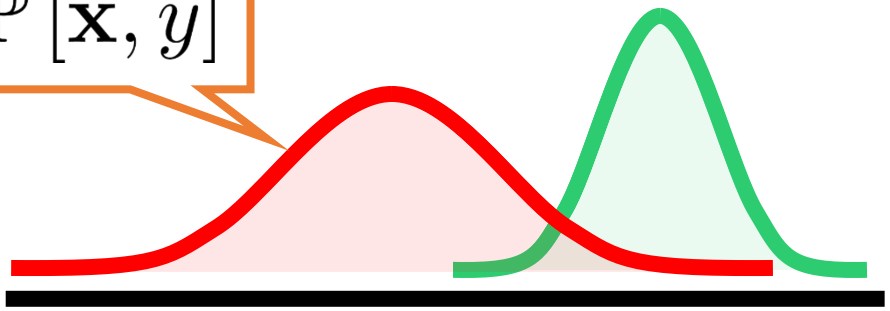
$$\mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{green}{\bullet}$$





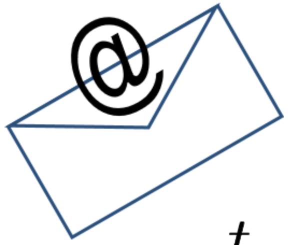
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label

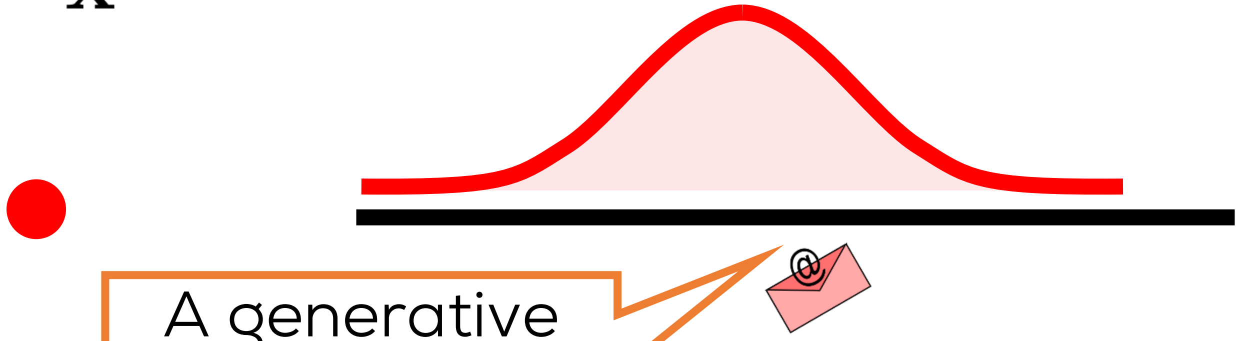


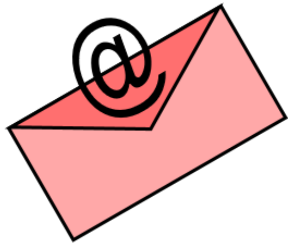
$\mathbf{x}^t$

$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{red}{\bullet}$$

$$\mathbb{P}\left[\textcolor{green}{\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[\textcolor{red}{\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = \textcolor{green}{\bullet}$$

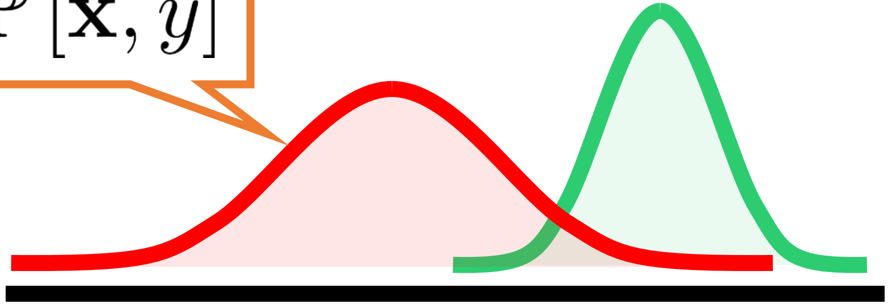




A generative model can generate data!

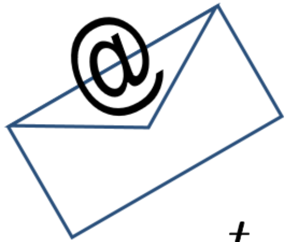
The generative story for labelled data

$\mathbb{P}\left[\mathbf{x}, y\right]$



A generative model can make predictions!

Predict the most likely label

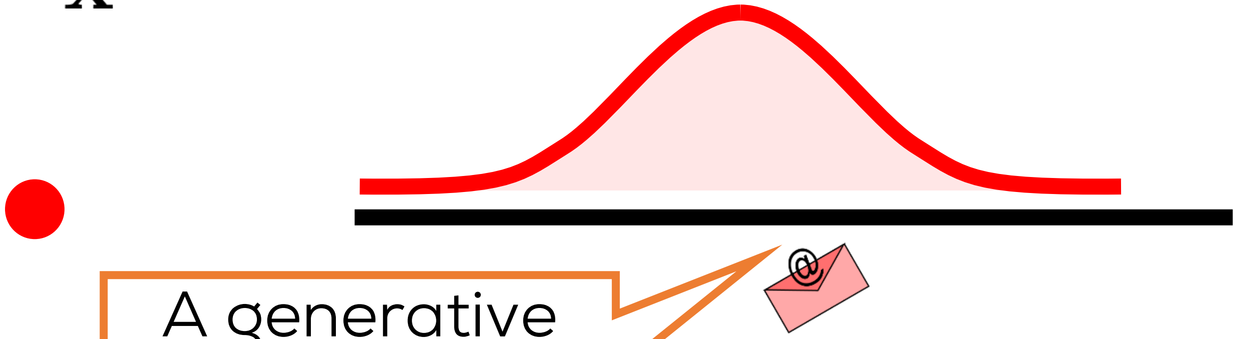


$\mathbf{x}^t$

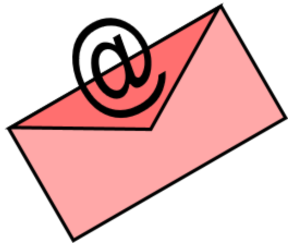
$$\mathbb{P}\left[y \,|\, \mathbf{x}^t\right] = \frac{\mathbb{P}\left[\mathbf{x}^t, y\right]}{\mathbb{P}\left[\mathbf{x}^t\right]}$$

$$\mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{red}\bullet}$$

$$\mathbb{P}\left[{\color{green}\bullet}, \mathbf{x}^t\right] > \mathbb{P}\left[{\color{red}\bullet}, \mathbf{x}^t\right] \quad \hat{y}^t = {\color{green}\bullet}$$



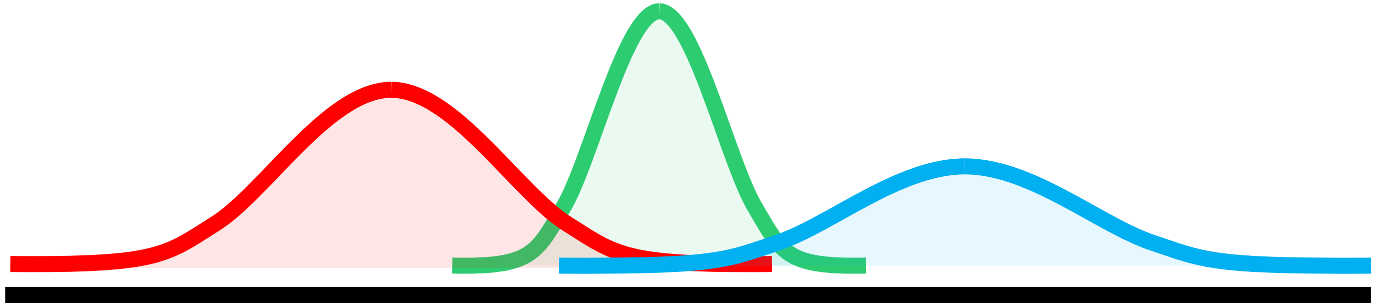
A generative model can generate data!



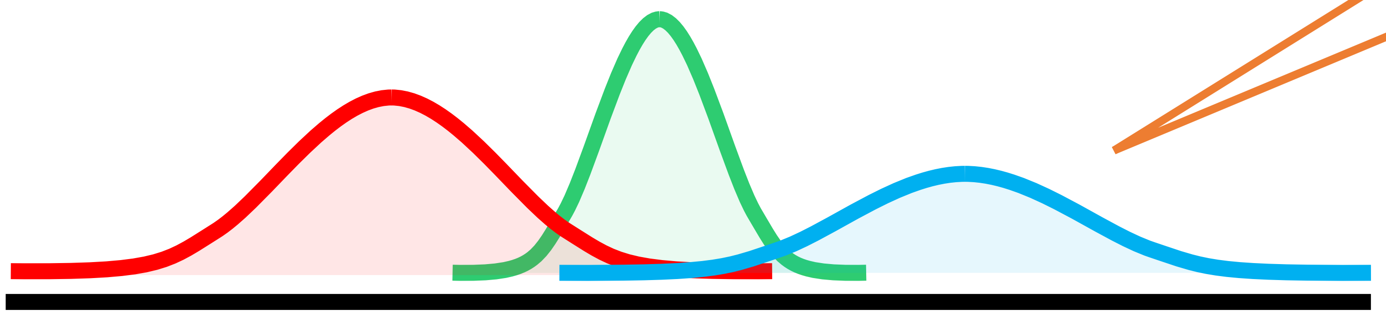
torch.ch

My First Generative Model

My First Generative Model

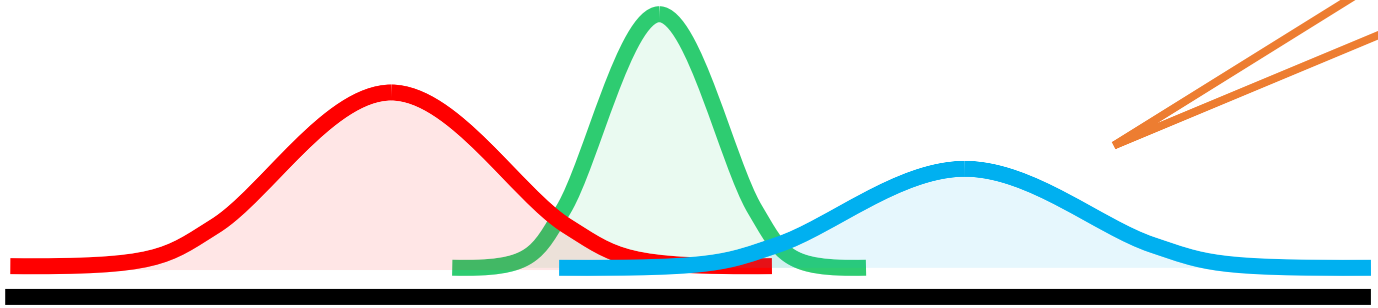


My First Generative Model



“Mixture” of
3 Gaussians

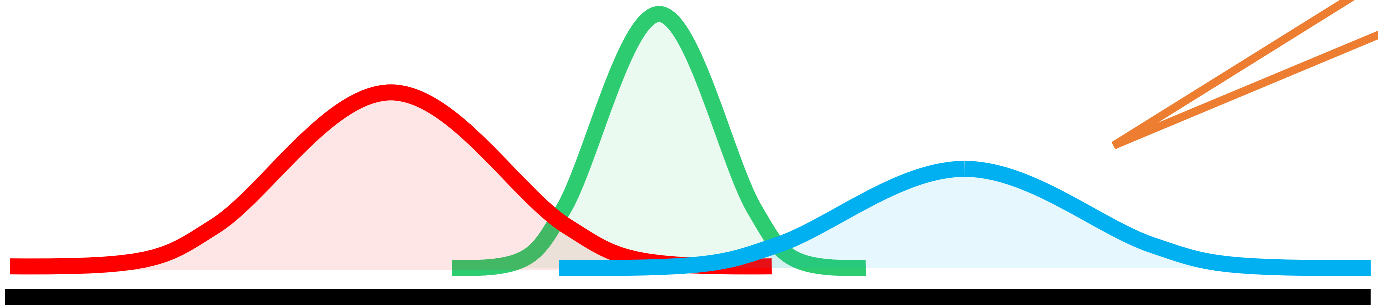
My First Generative Model



"Mixture" of 3 Gaussians

On the real line

My First Generative Model

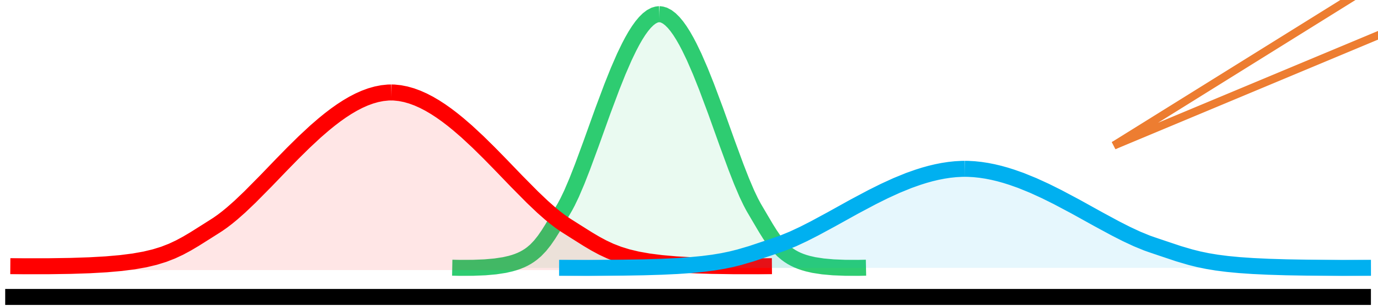


“Mixture” of 3 Gaussians

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

My First Generative Model



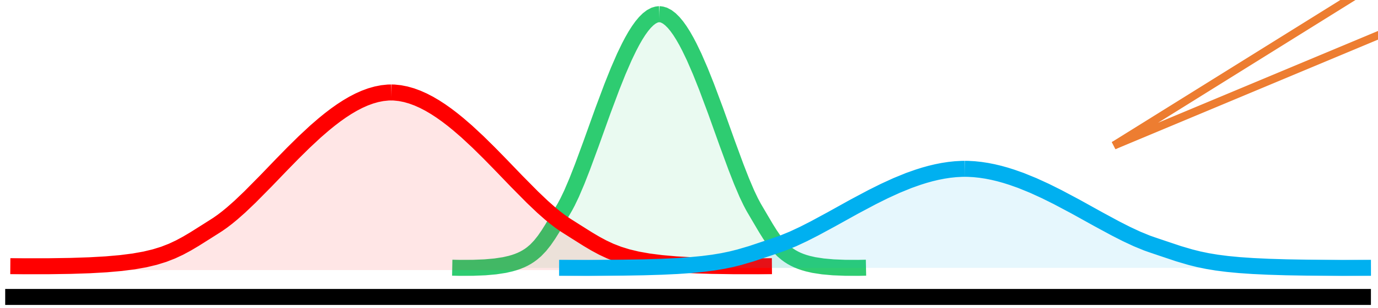
"Mixture" of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

My First Generative Model



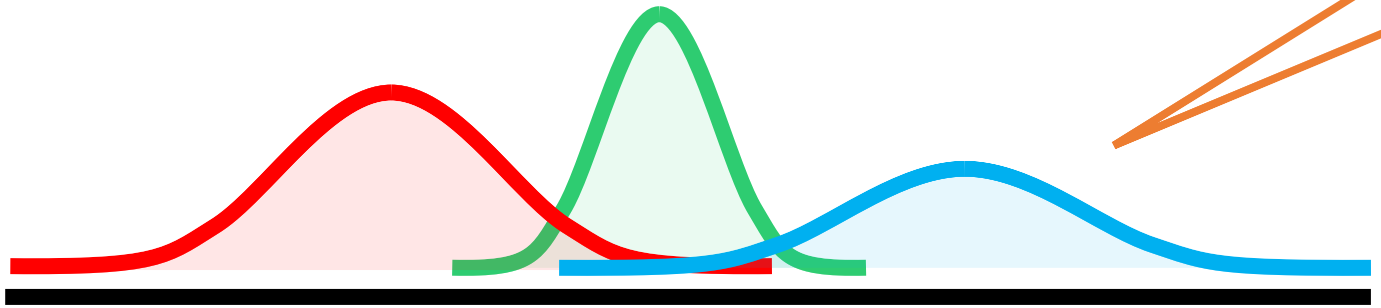
"Mixture" of 3 Gaussians

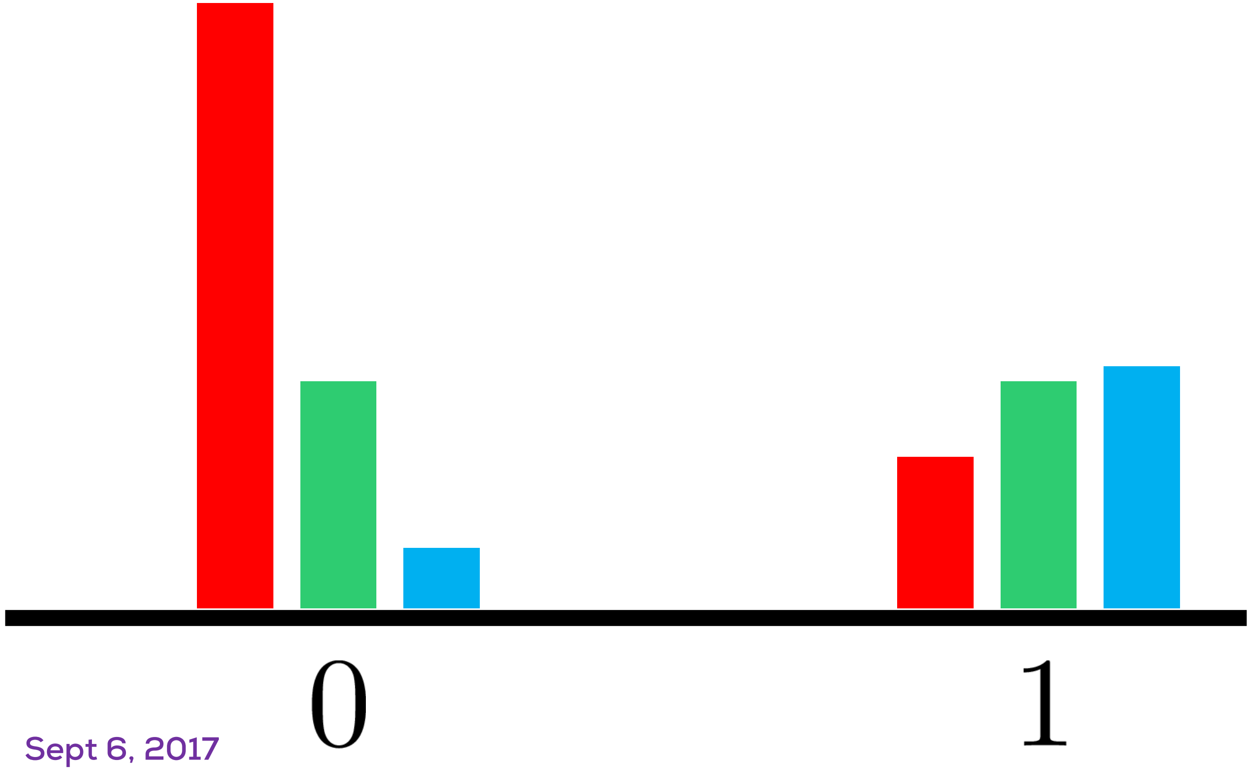
Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$\mathrm{y} \in \{1, 2, 3\}$

My First Generative Model





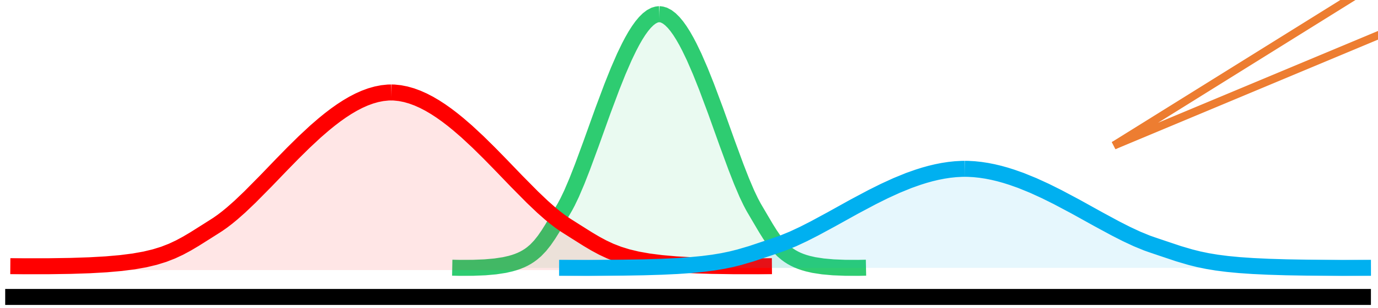
"Mixture" of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

My First Generative Model

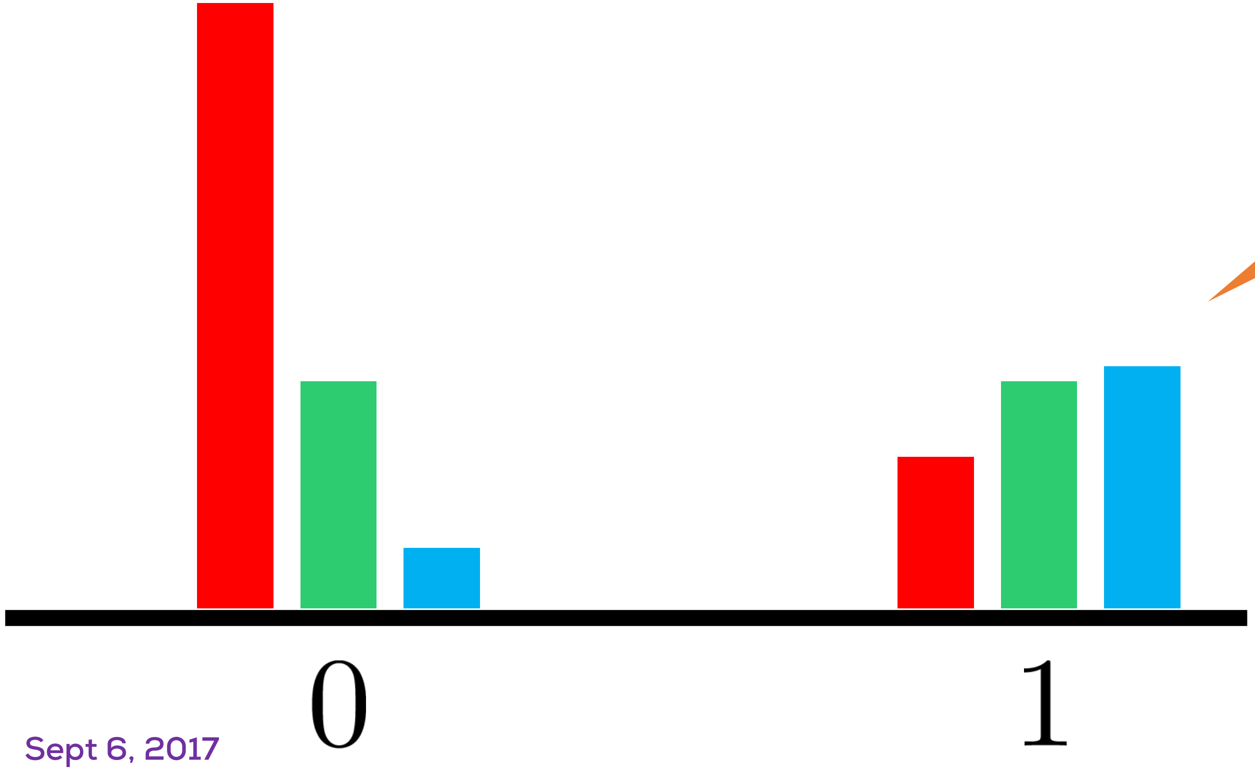


“Mixture” of 3 Gaussians

Can we try something simpler first?

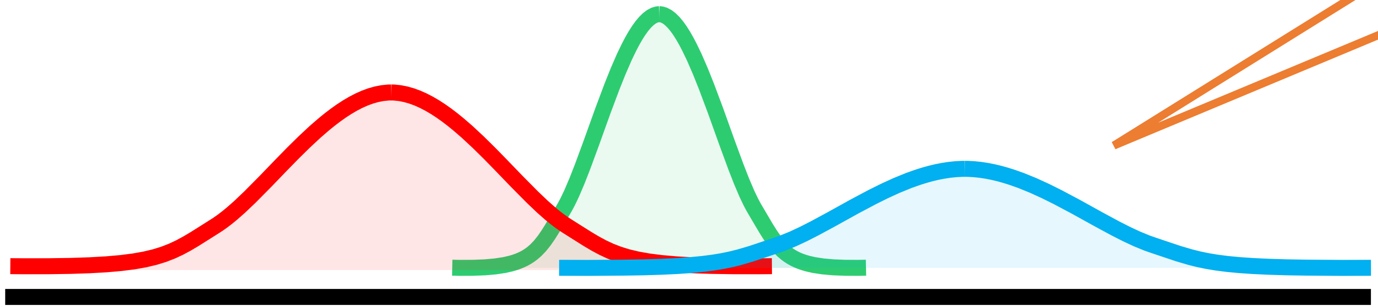
On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$\mathrm{y} \in \{1, 2, 3\}$



“Mixture” of 3 Bernoulli’s

My First Generative Model

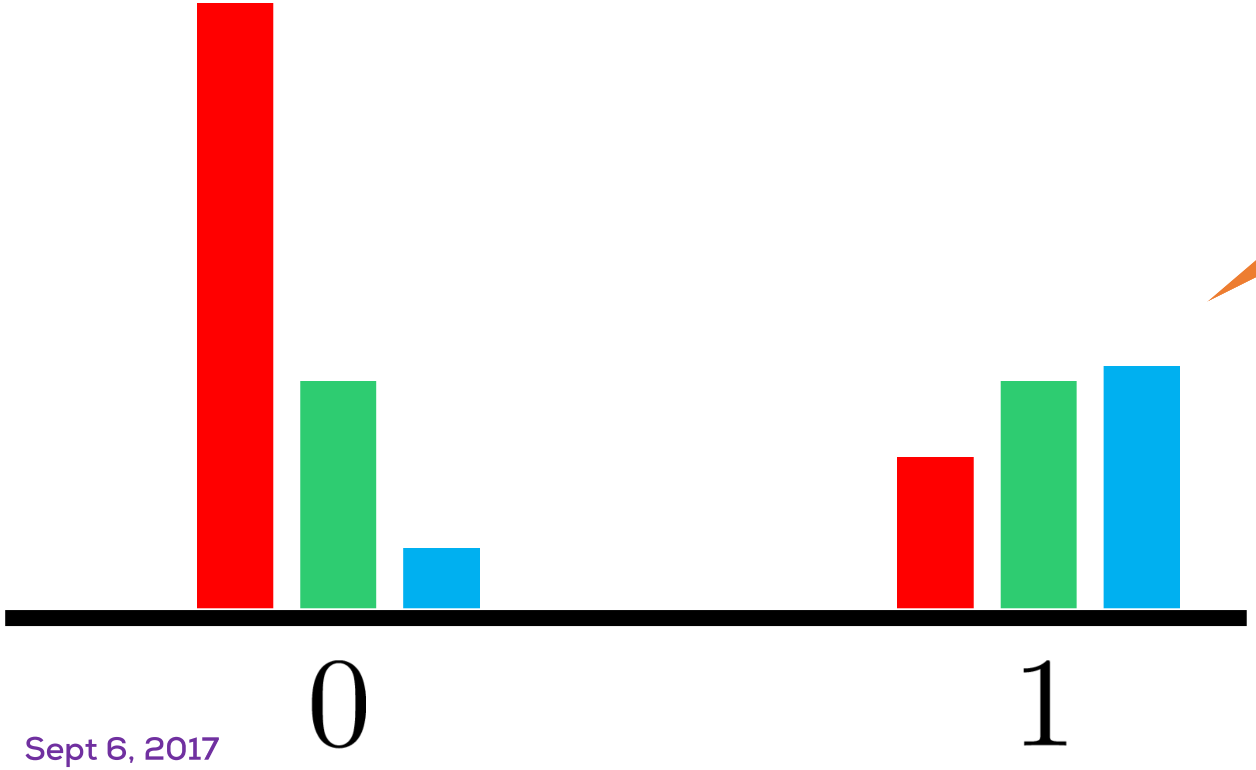


"Mixture" of 3 Gaussians

Can we try something simpler first?

On the real line

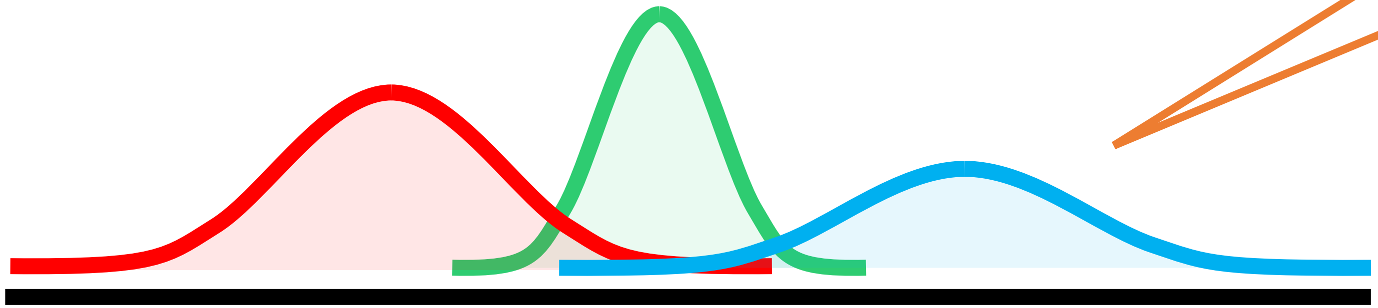
$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$



"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

My First Generative Model

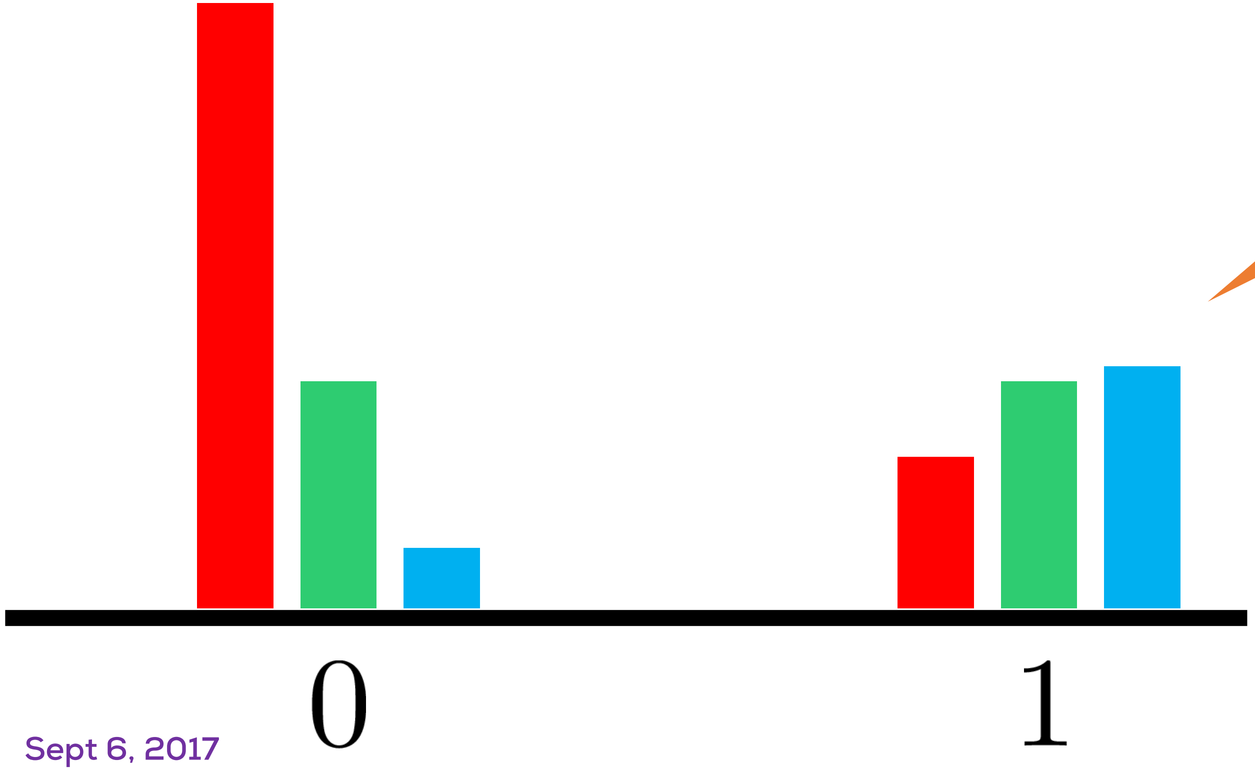


“Mixture” of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

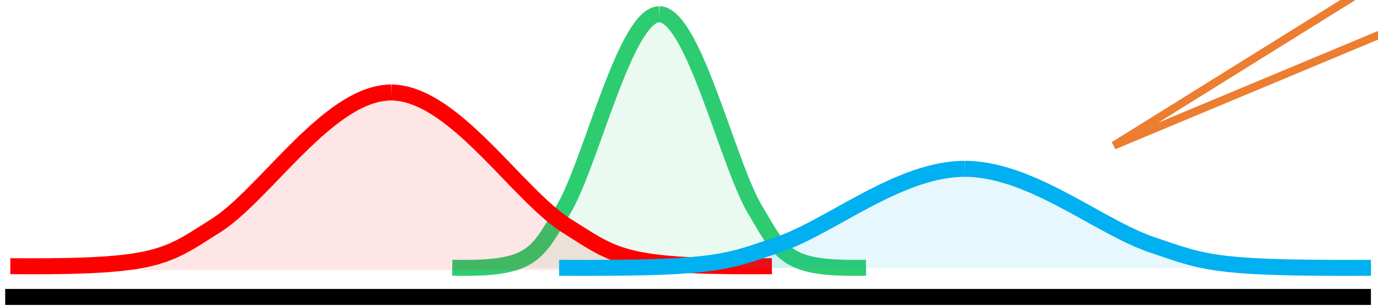


“Mixture” of 3 Bernoulli’s

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

E.g. x can denote whether the word “urgent” appears in the mail or not

My First Generative Model

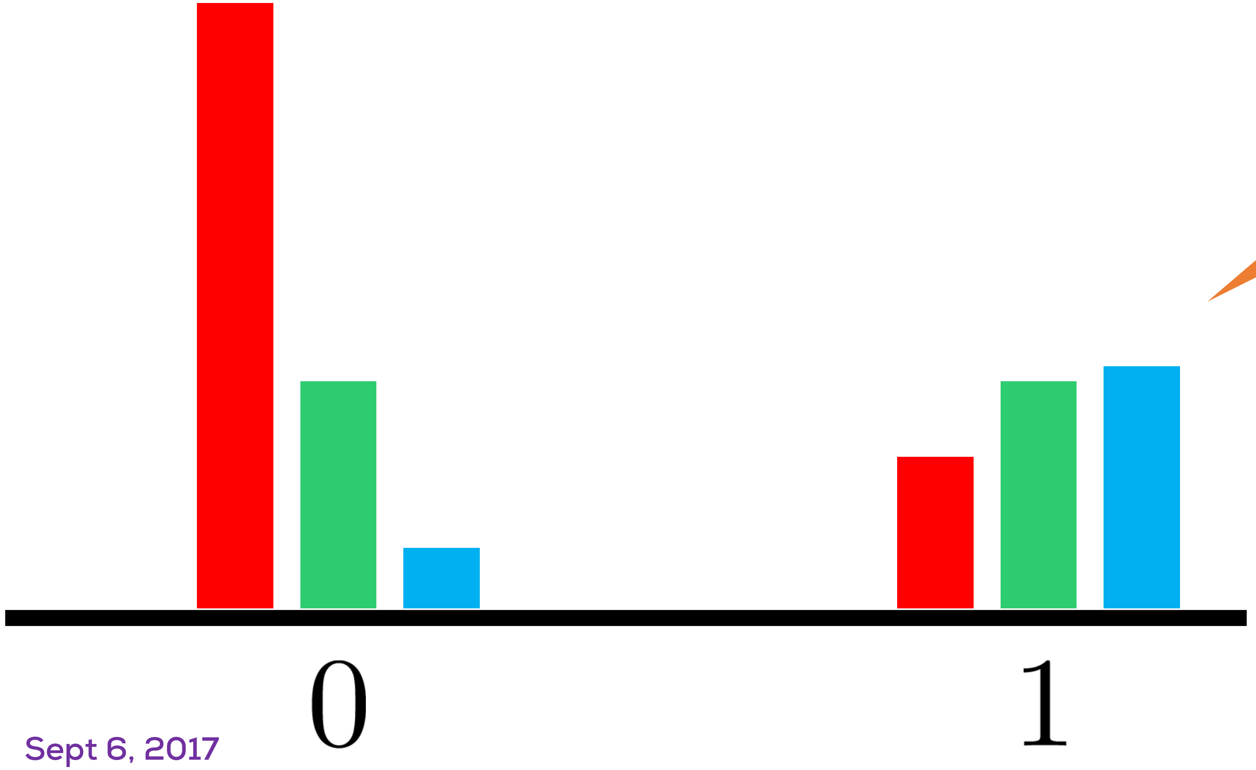


“Mixture” of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$



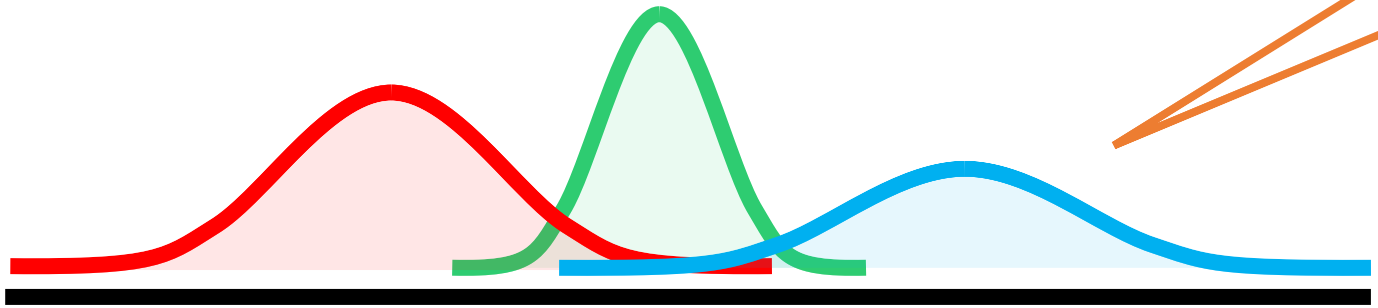
“Mixture” of 3 Bernoulli’s

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word “urgent” appears in the mail or not

My First Generative Model

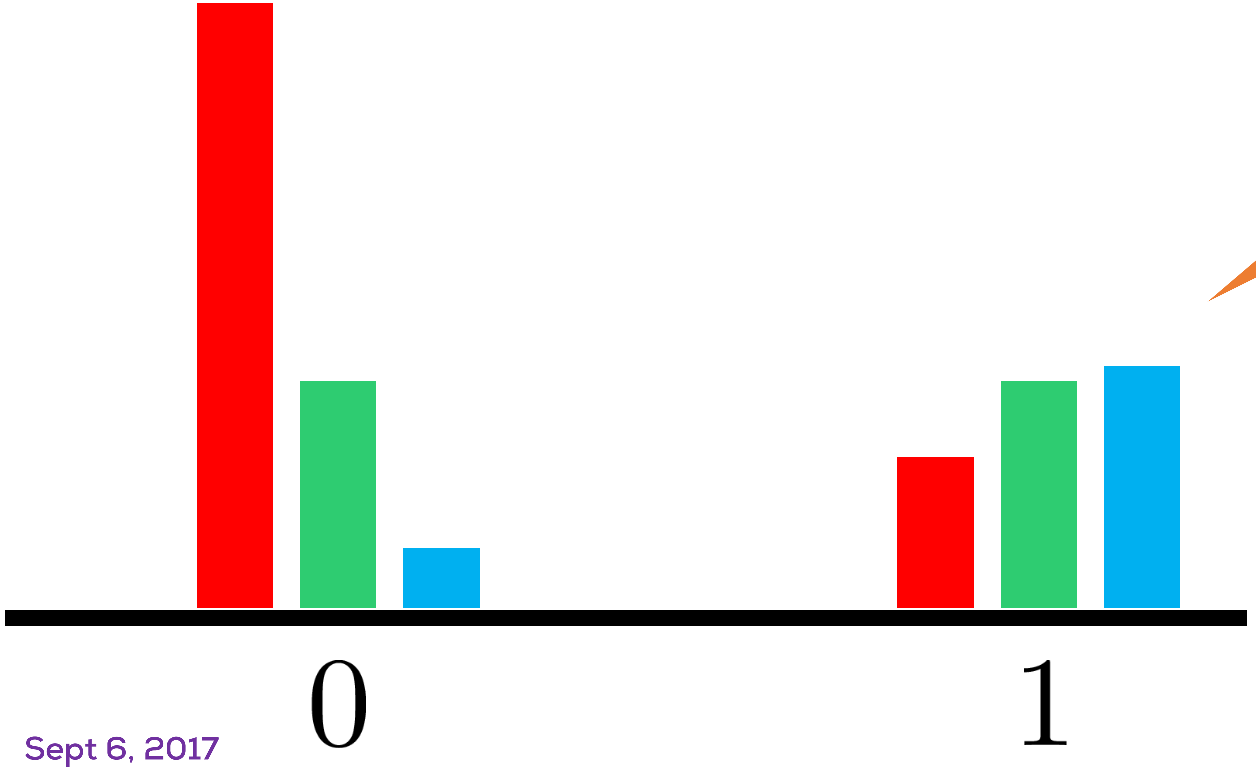


"Mixture" of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$



"Mixture" of 3 Bernoulli's

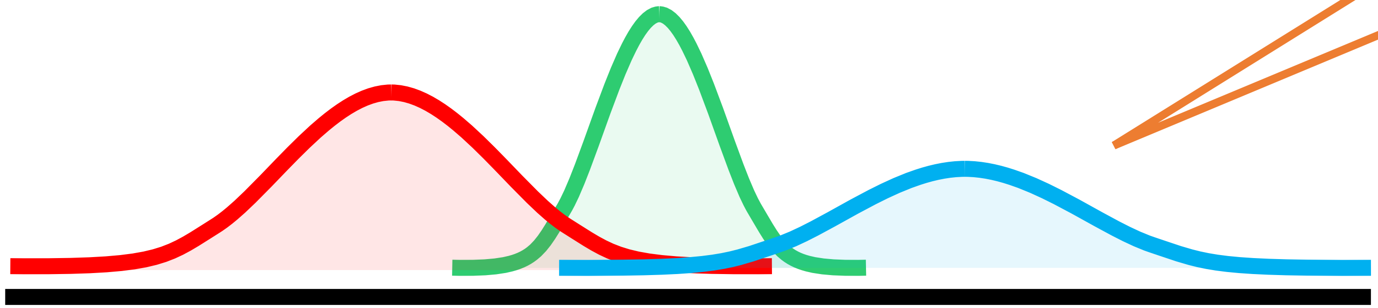
$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

Learning $\mathbb{P}[x|y]$ is key!

My First Generative Model

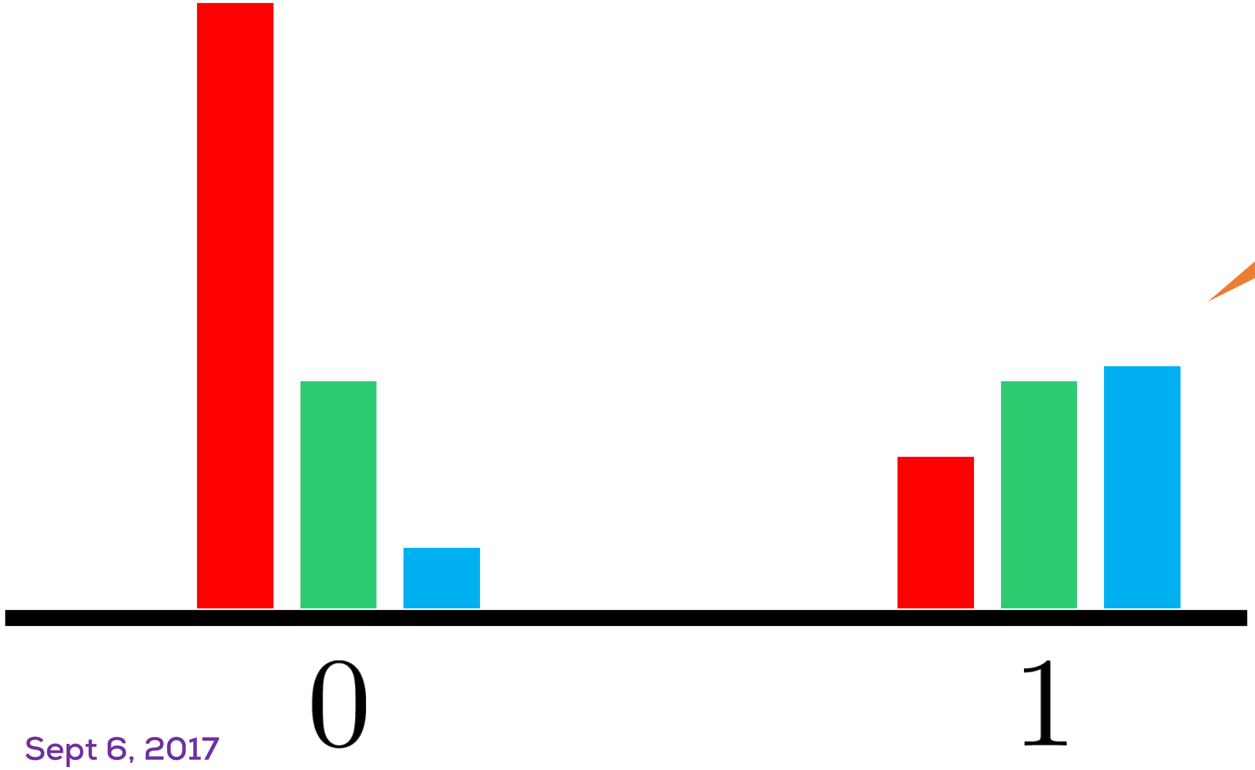


"Mixture" of 3 Gaussians

Can we try something simpler first?

On the real line

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$



"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

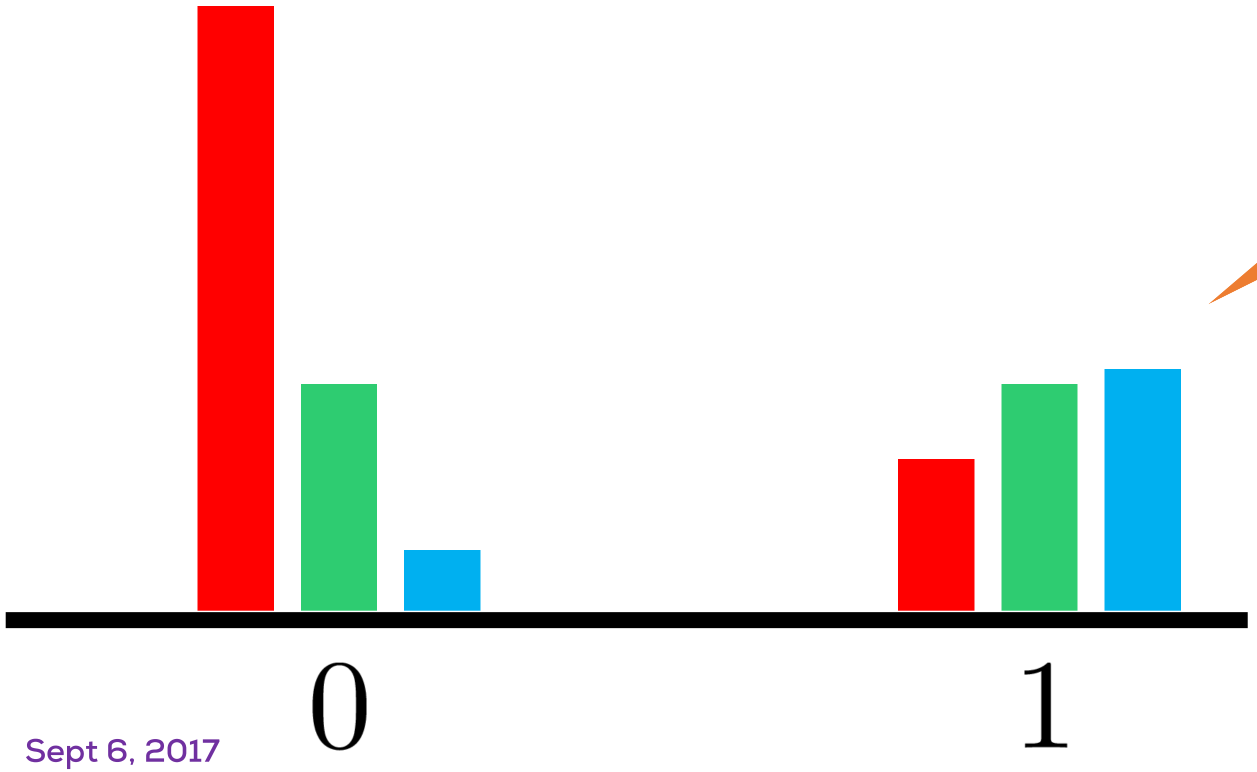
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

Learning $\mathbb{P}[x|y]$ is key!

$\mathbb{P}[x, y]$
$= \mathbb{P}[x|y] \cdot \mathbb{P}[y]$

My First Generative Model



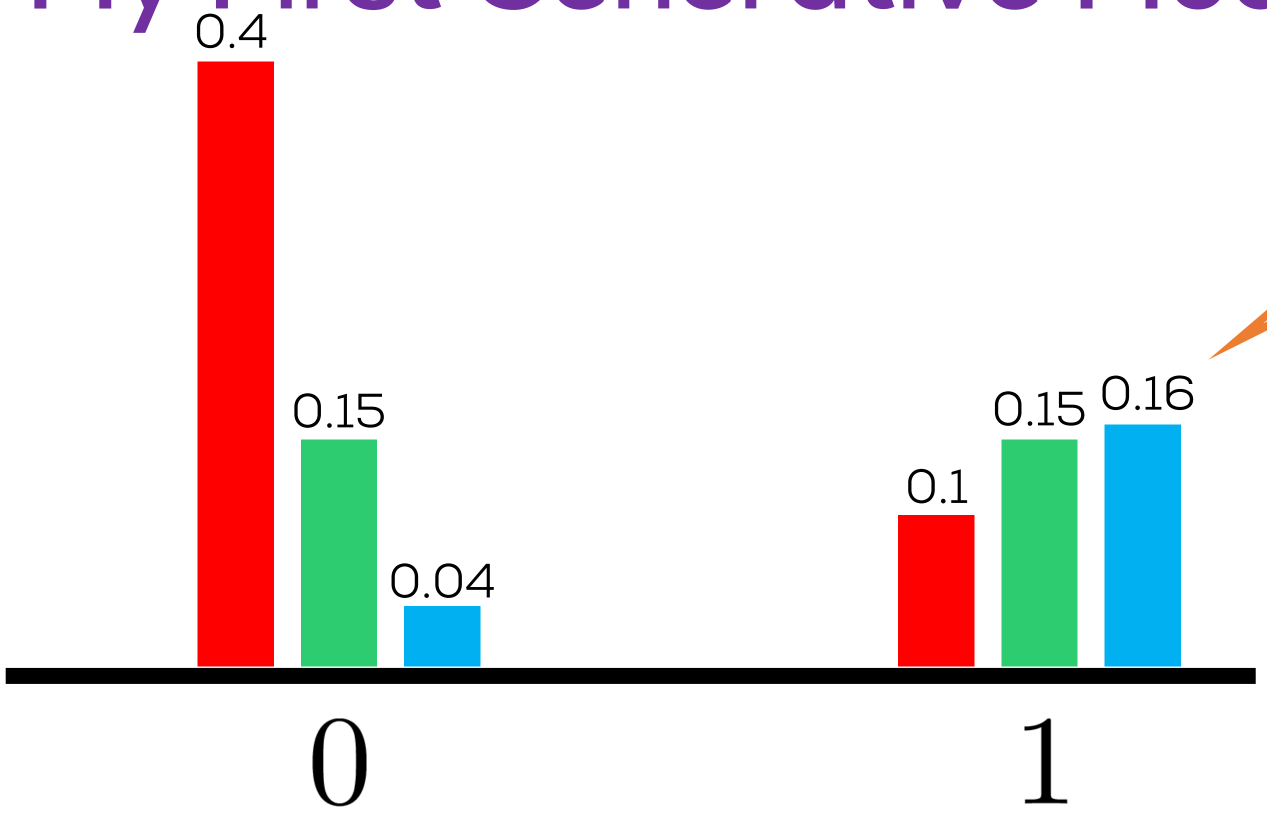
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

My First Generative Model



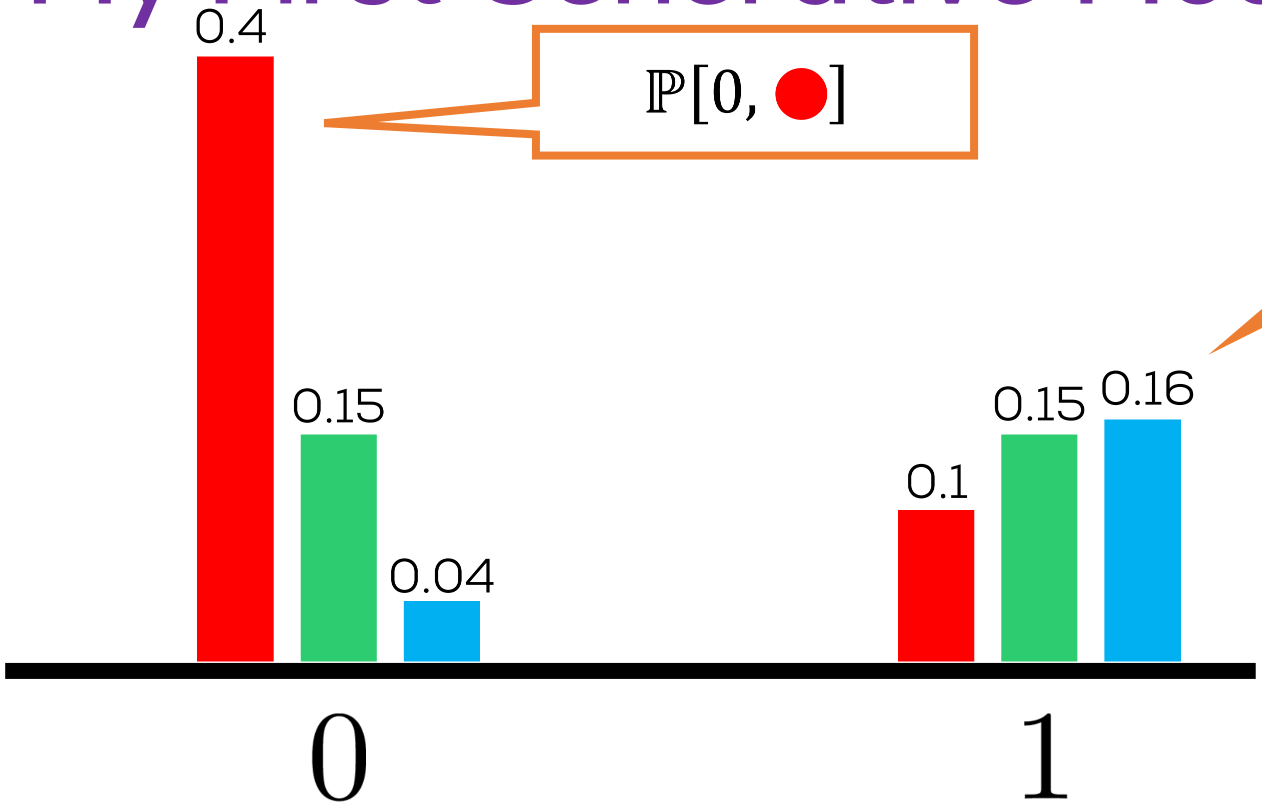
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$, $y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

My First Generative Model



$\mathbb{P}[0, \bullet]$

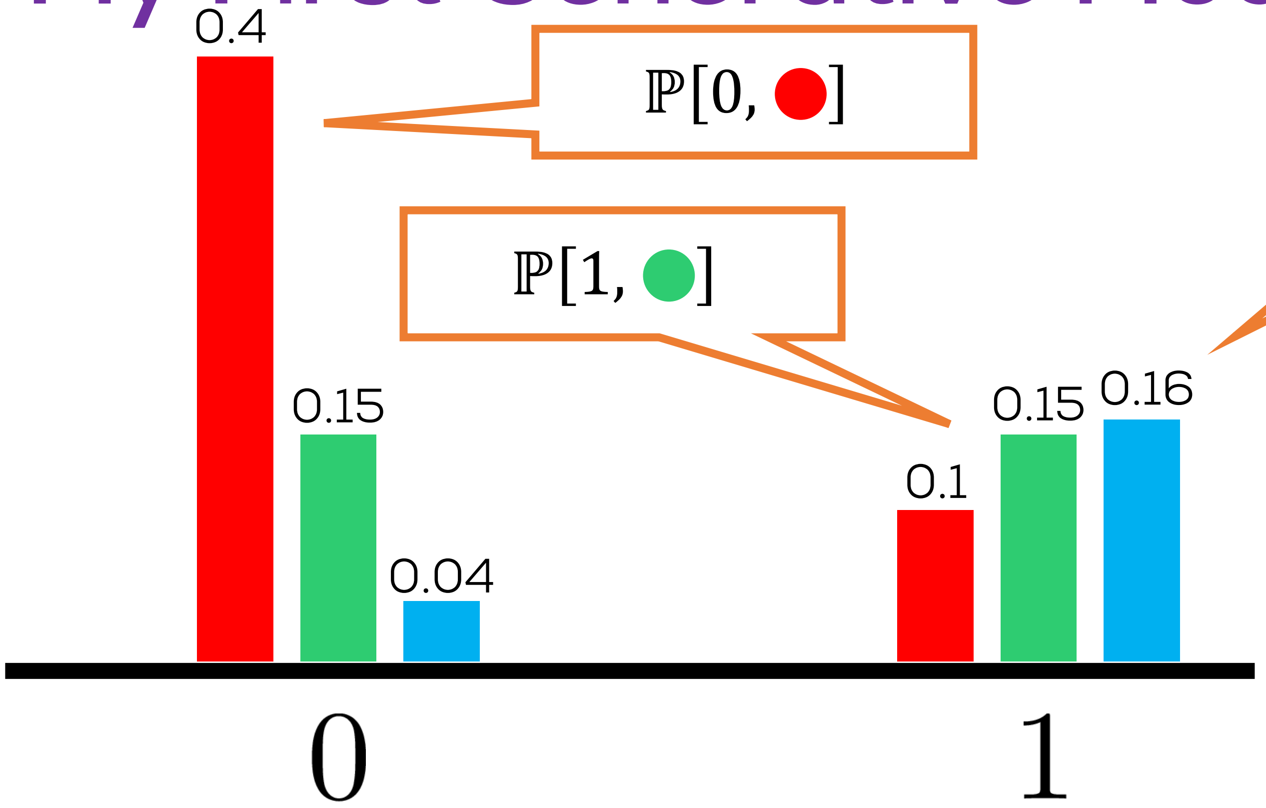
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

My First Generative Model



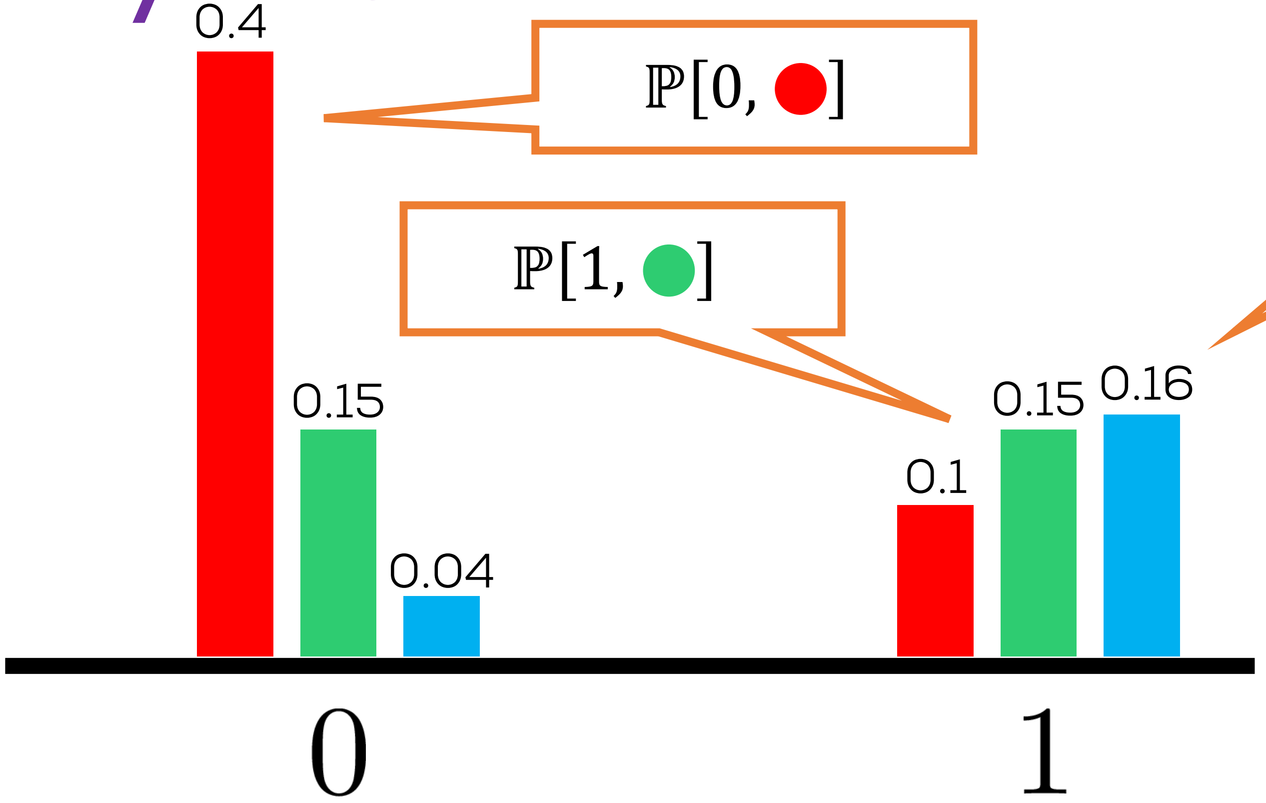
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

My First Generative Model

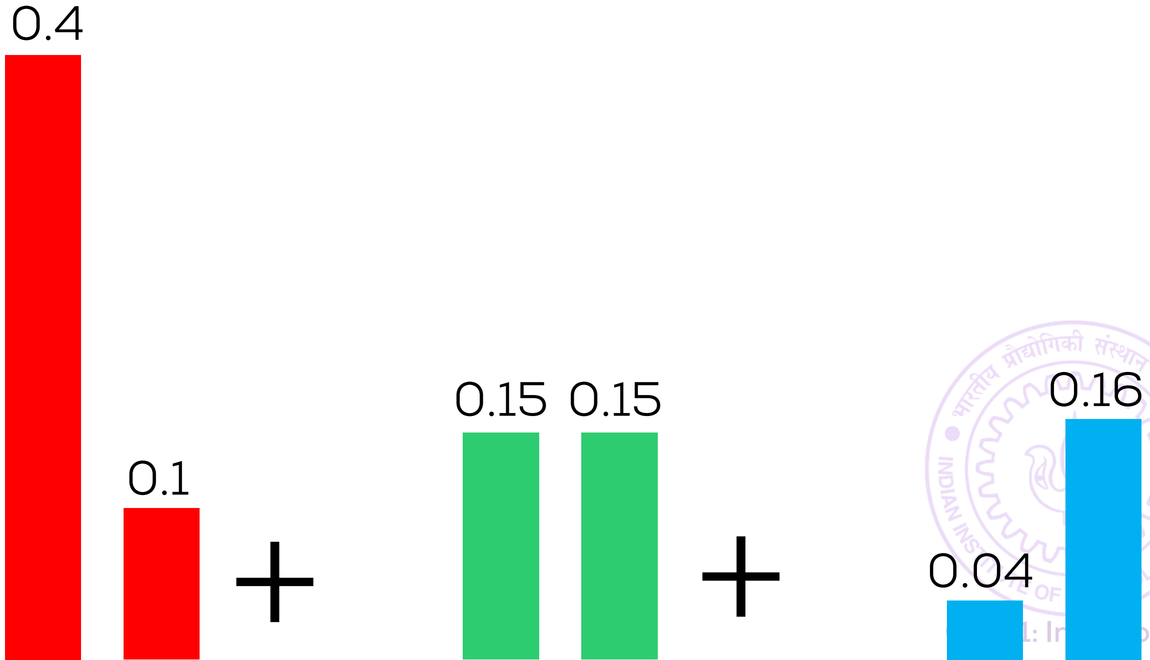


"Mixture" of 3 Bernoulli's

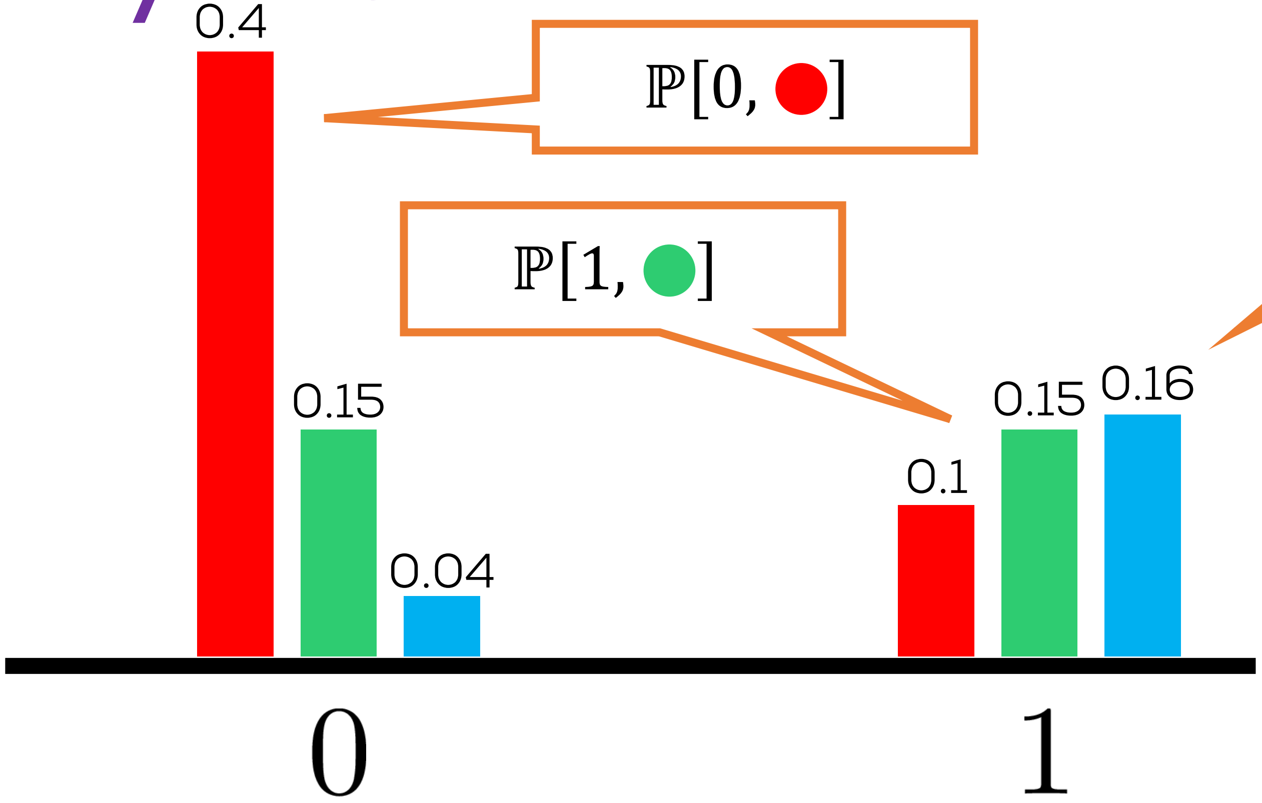
$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not



My First Generative Model



$\mathbb{P}[0, \text{●}]$

$\mathbb{P}[1, \text{●}]$

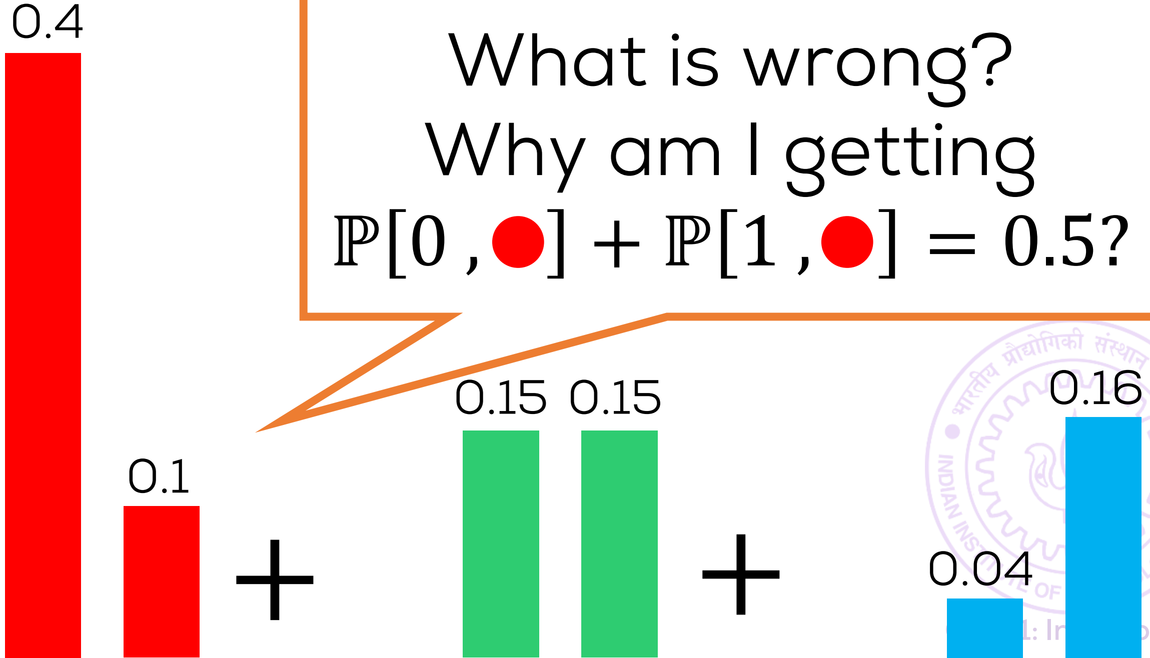
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

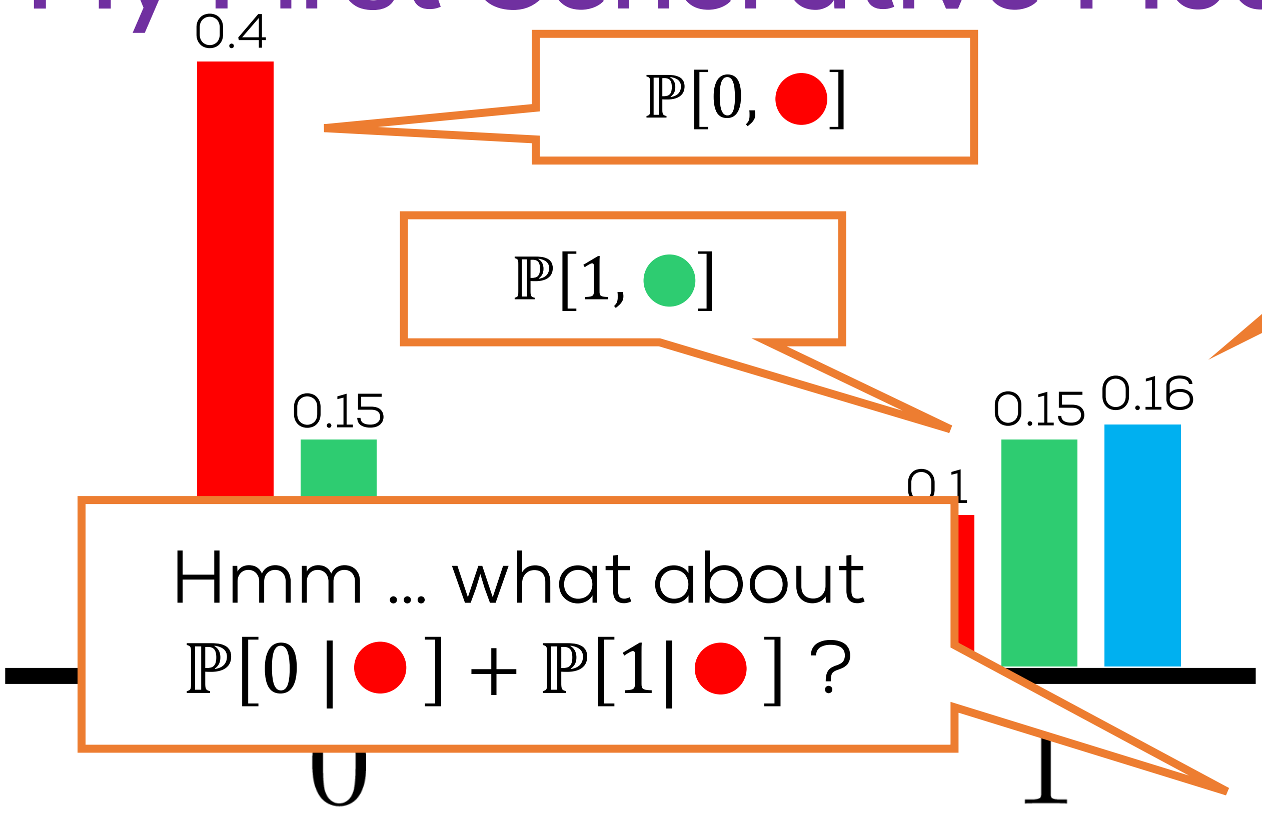
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

What is wrong? Why am I getting $\mathbb{P}[0, \text{●}] + \mathbb{P}[1, \text{●}] = 0.5$?



My First Generative Model



$\mathbb{P}[0, ●]$

$\mathbb{P}[1, ●]$

Hmm … what about $\mathbb{P}[0\,|\,●] + \mathbb{P}[1|\,●]$?

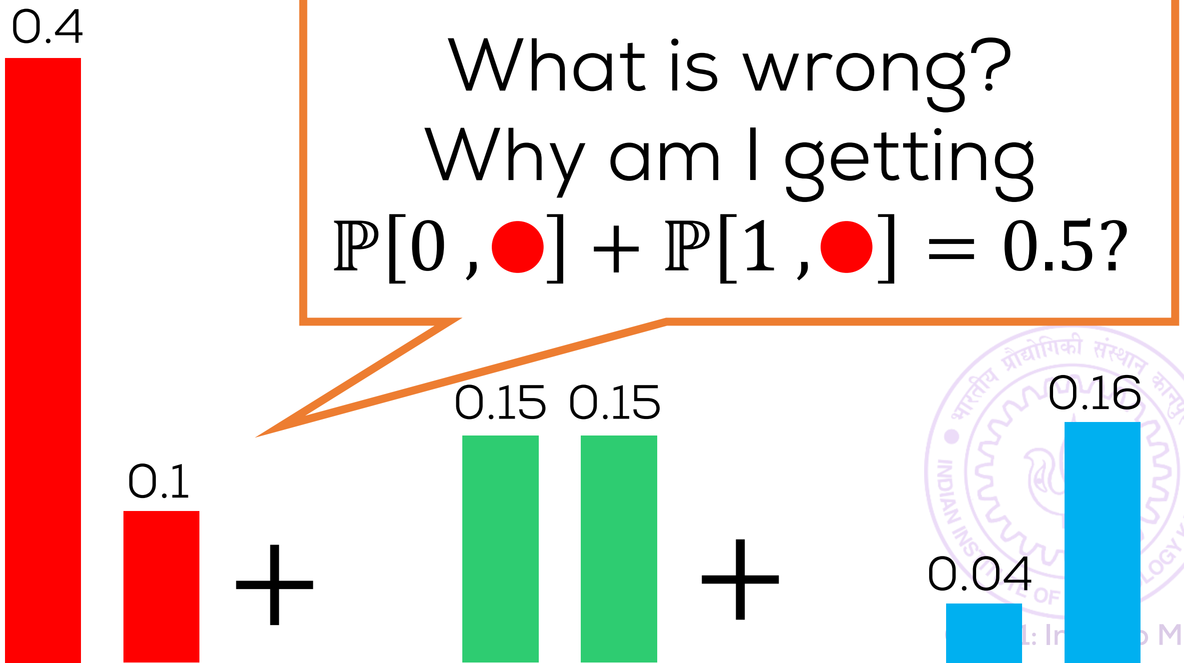
"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

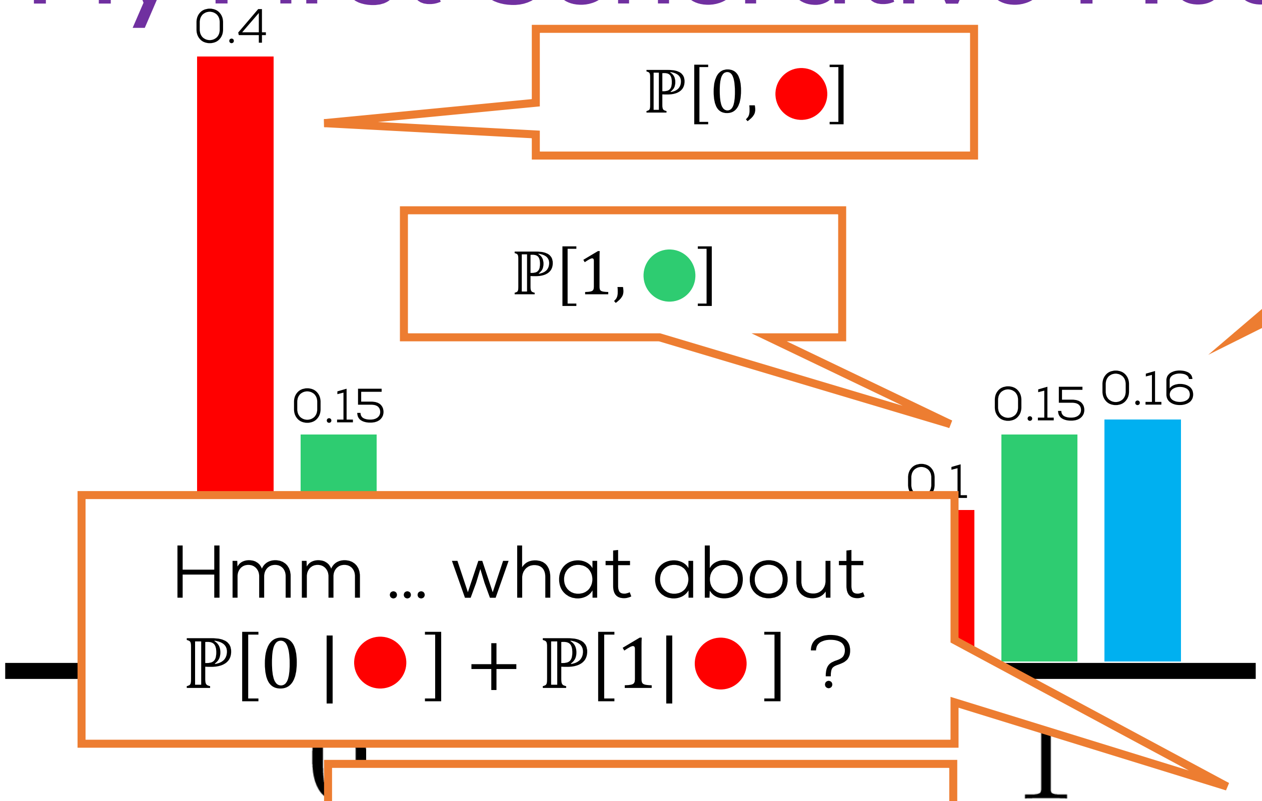
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

What is wrong?
Why am I getting
$\mathbb{P}[0\,, ●] + \mathbb{P}[1\,, ●] = 0.5$?



My First Generative Model



$\mathbb{P}[0, \bullet]$

$\mathbb{P}[1, \bullet]$

"Mixture" of 3 Bernoulli's

$x \in \{0,1\}$,
$y \in \{1, 2, 3\}$

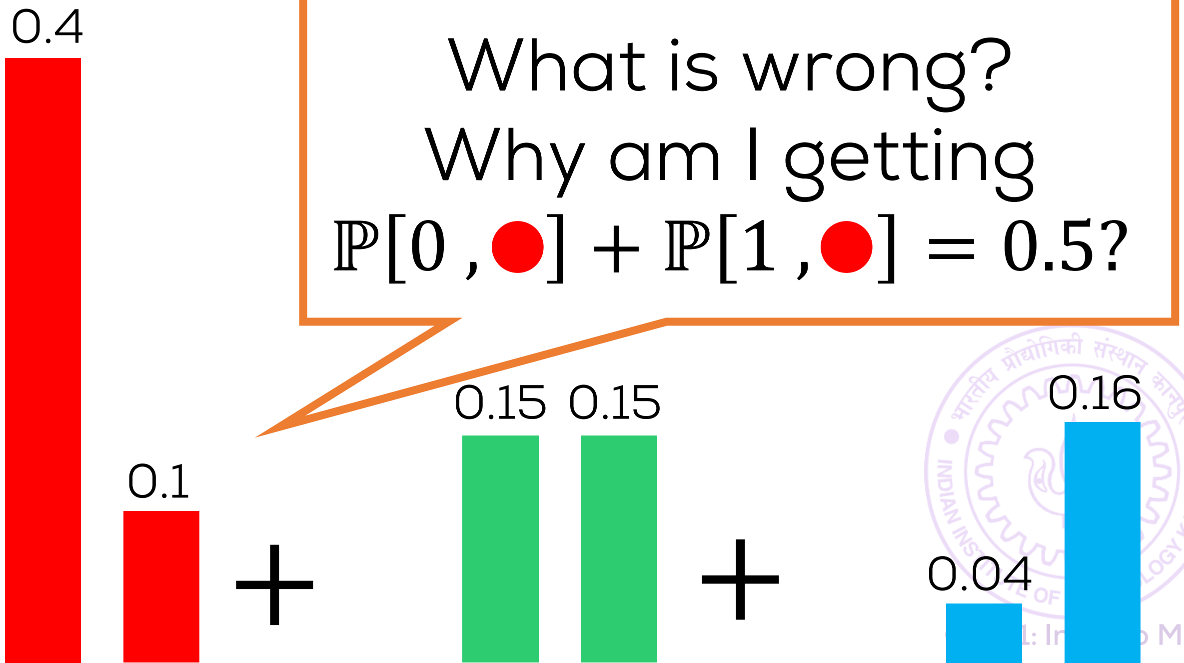
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

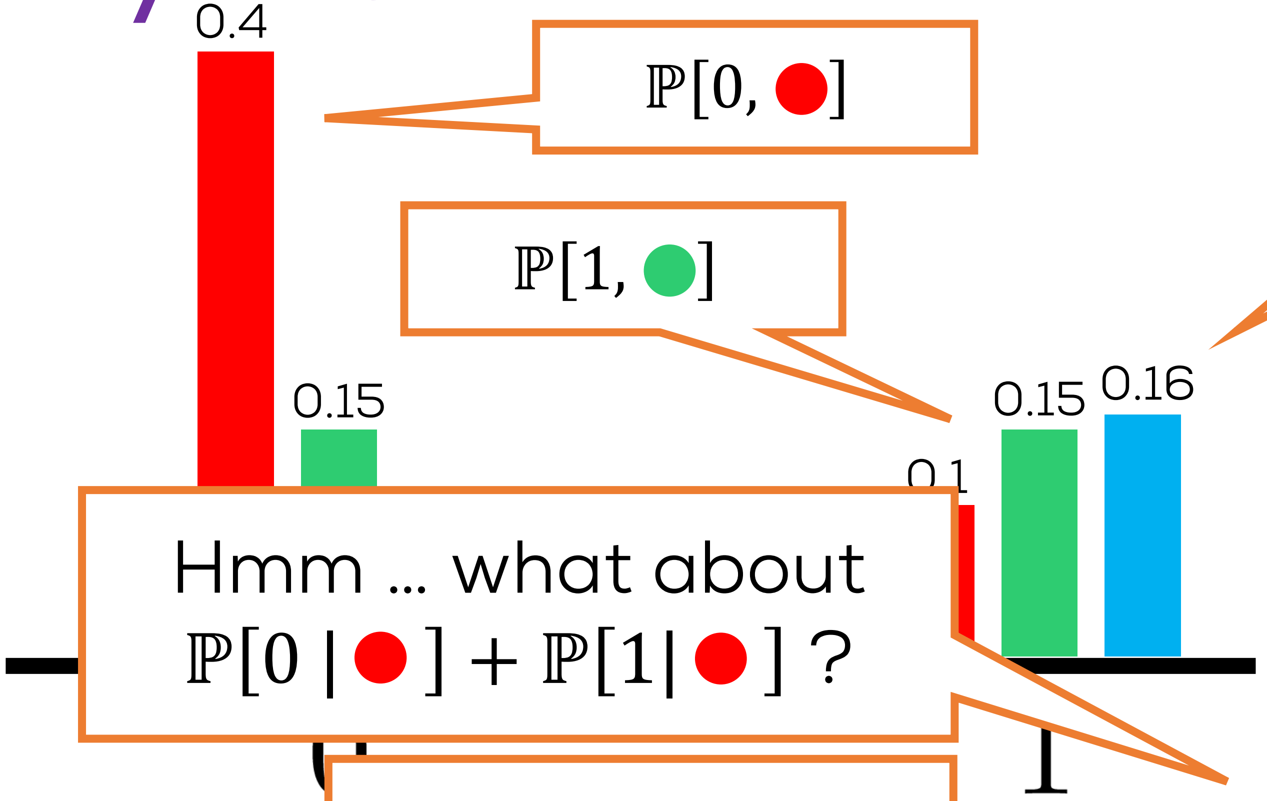
Hmm … what about $\mathbb{P}[0 \mid \bullet] + \mathbb{P}[1 \mid \bullet]$?

$$\mathbb{P}[0 \mid \bullet] = \frac{\mathbb{P}[0, \bullet]}{\mathbb{P}[\bullet]}$$

What is wrong? Why am I getting $\mathbb{P}[0, \bullet] + \mathbb{P}[1, \bullet] = 0.5$?



My First Generative Model



$\mathbb{P}[0, \color{red}{\bullet}]$

$\mathbb{P}[1, \color{green}{\bullet}]$

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{\color{red}{1}, \color{green}{2}, \color{cyan}{3}\}$

Can I learn this mixture?

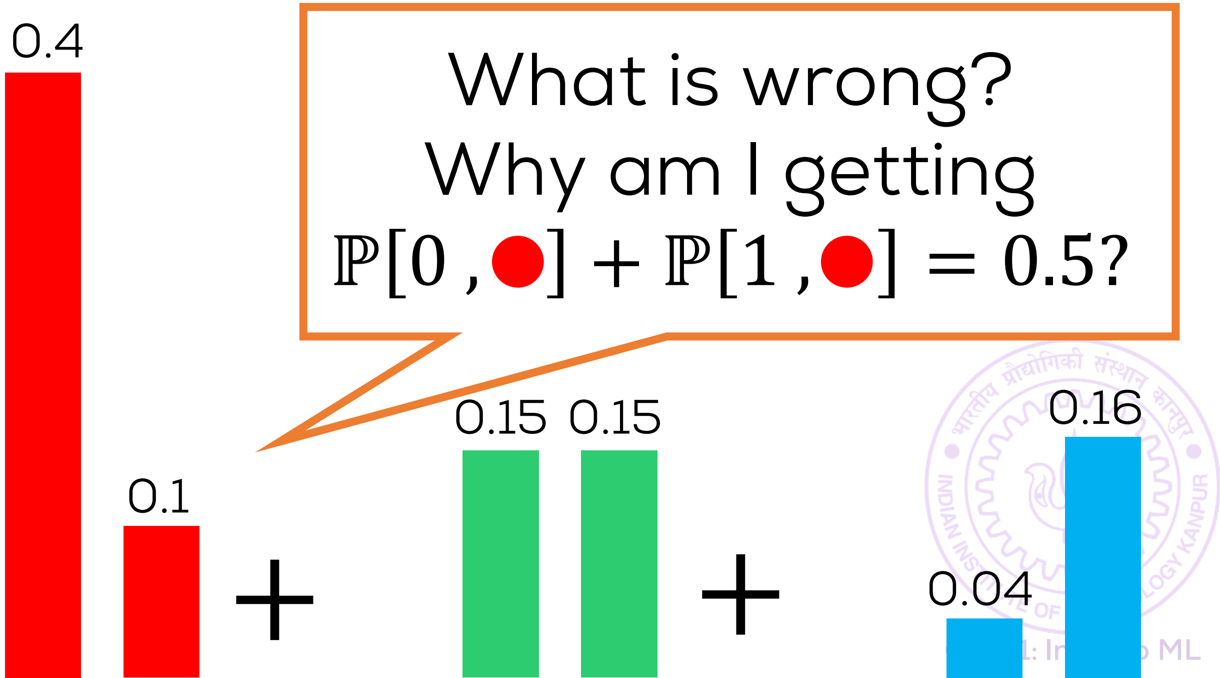
E.g. x can denote whether the word "urgent" appears in the mail or not

Hmm … what about $\mathbb{P}[0\,|\,\color{red}{\bullet}] + \mathbb{P}[1\,|\,\color{red}{\bullet}]$?

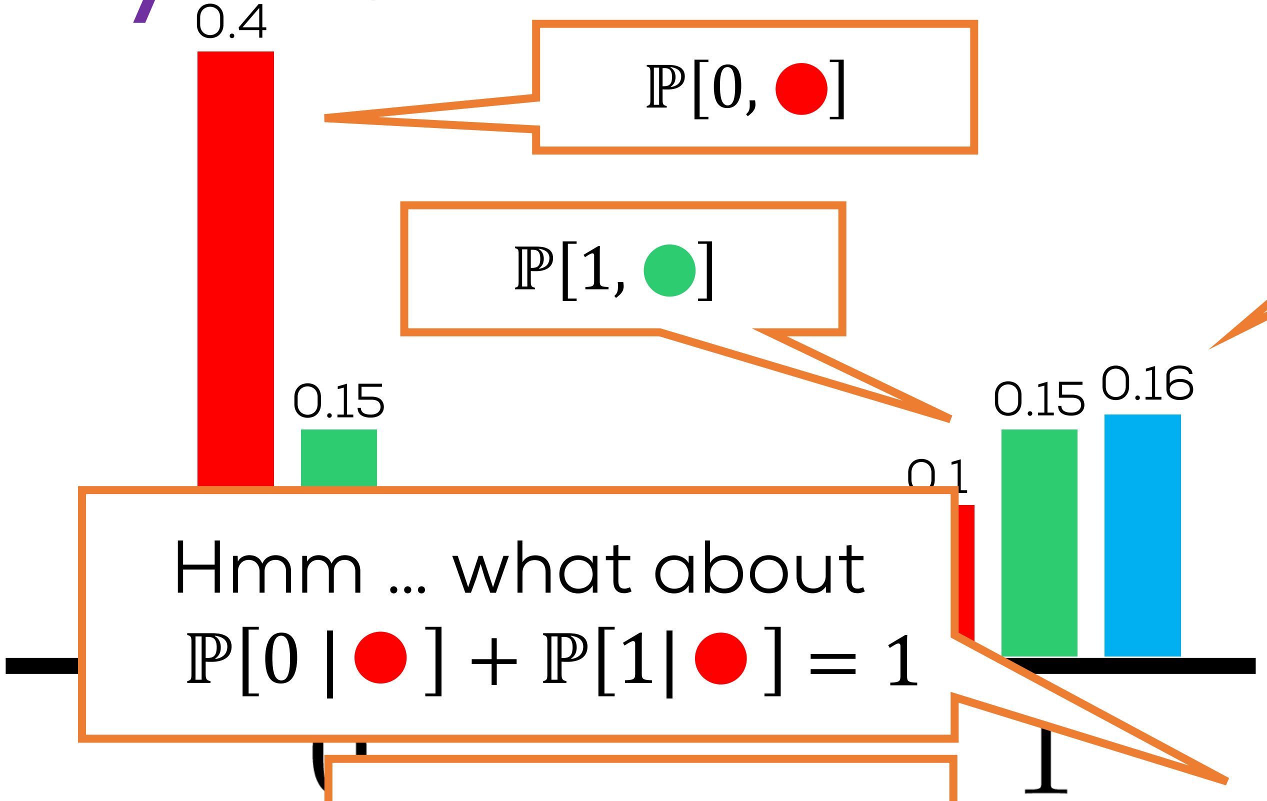
$$\mathbb{P}[0\,|\,\color{red}{\bullet}] = \frac{\mathbb{P}[0, \color{red}{\bullet}]}{\mathbb{P}[\color{red}{\bullet}]}$$

$$\mathbb{P}[\color{red}{\bullet}] = \mathbb{P}[0\,, \color{red}{\bullet}] + \mathbb{P}[1\,, \color{red}{\bullet}]$$

What is wrong? Why am I getting $\mathbb{P}[0\,, \color{red}{\bullet}] + \mathbb{P}[1\,, \color{red}{\bullet}] = 0.5$?



My First Generative Model



$\mathbb{P}[0, ●]$

$\mathbb{P}[1, ●]$

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

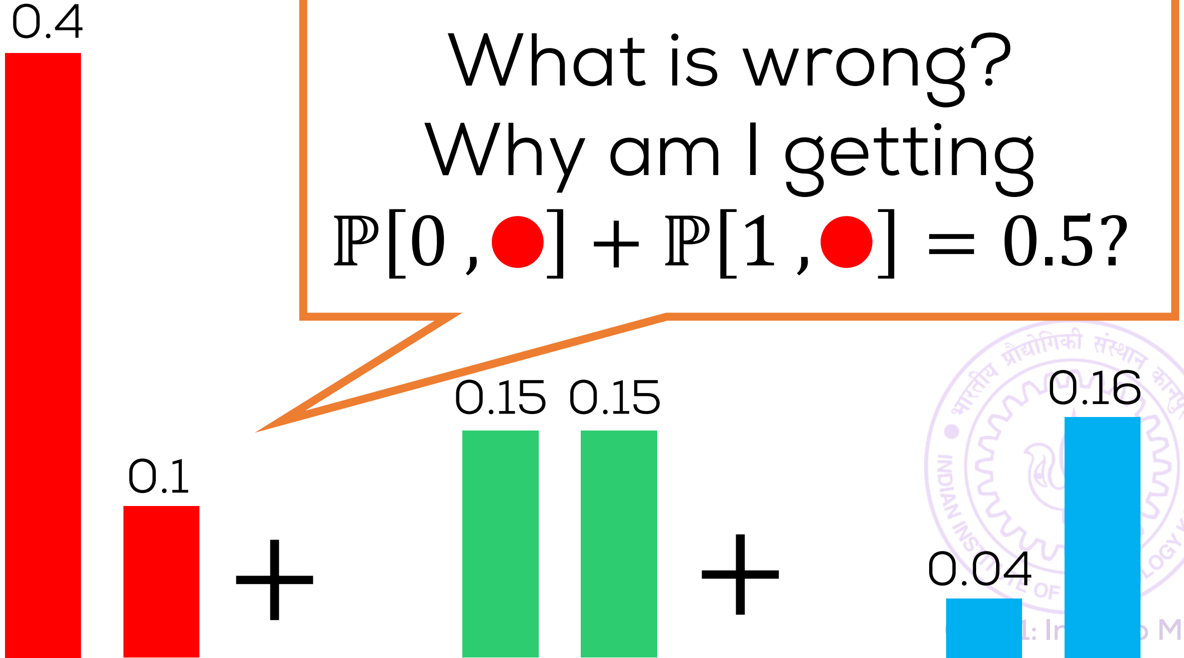
E.g. x can denote whether the word "urgent" appears in the mail or not

Hmm … what about
$\mathbb{P}[0 \mid ●] + \mathbb{P}[1 \mid ●] = 1$

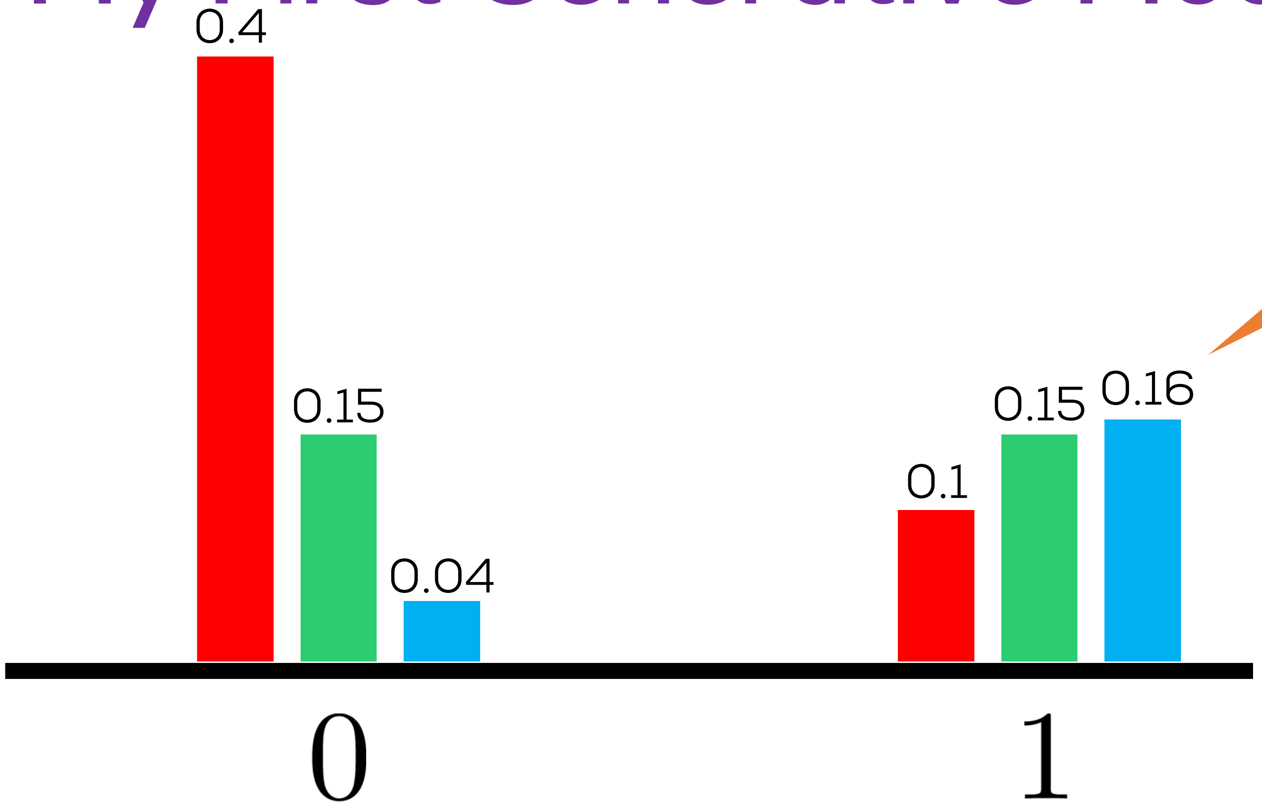
$$\mathbb{P}[0 \mid ●] = \frac{\mathbb{P}[0, ●]}{\mathbb{P}[●]}$$

$$\mathbb{P}[●] = \mathbb{P}[0, ●] + \mathbb{P}[1, ●]$$

What is wrong?
Why am I getting
$\mathbb{P}[0, ●] + \mathbb{P}[1, ●] = 0.5$?



My First Generative Model

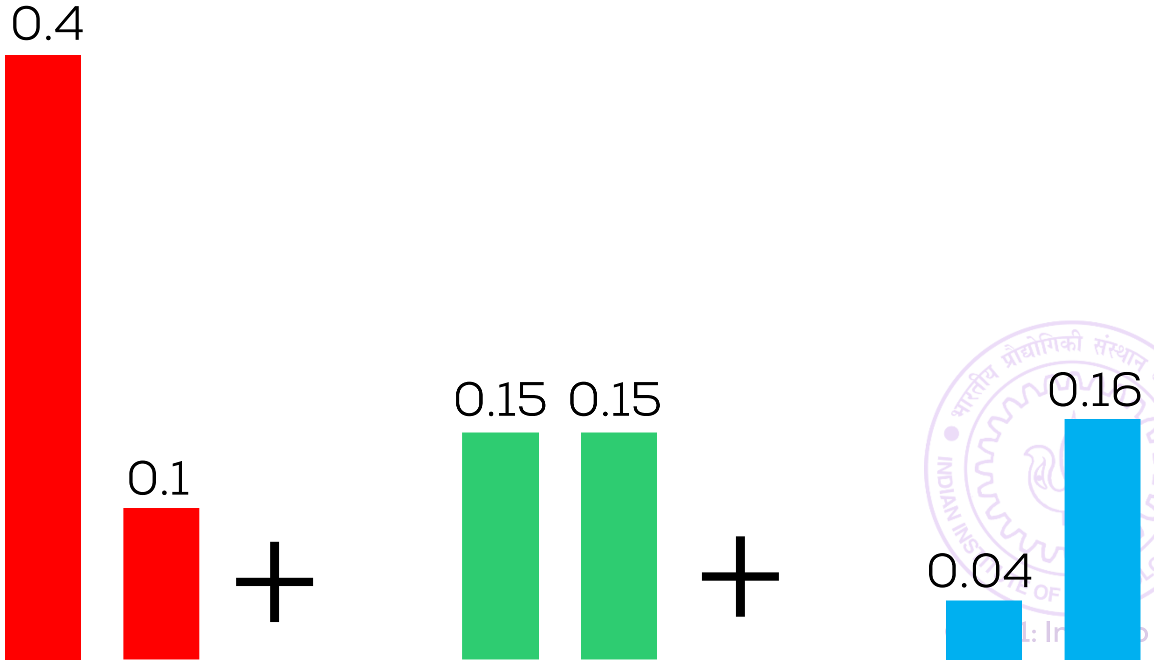


"Mixture" of 3 Bernoulli's

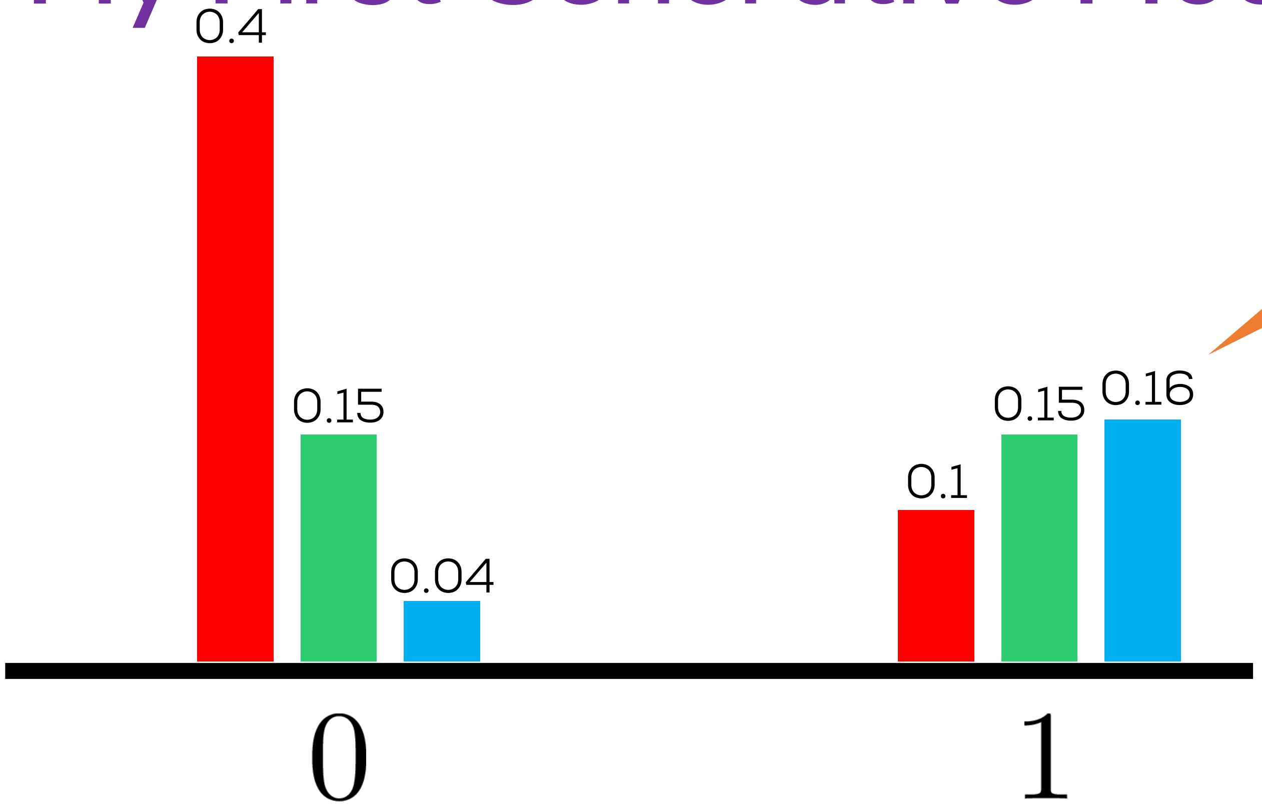
$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{\color{red}{1}, \color{green}{2}, \color{cyan}{3}\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not



My First Generative Model

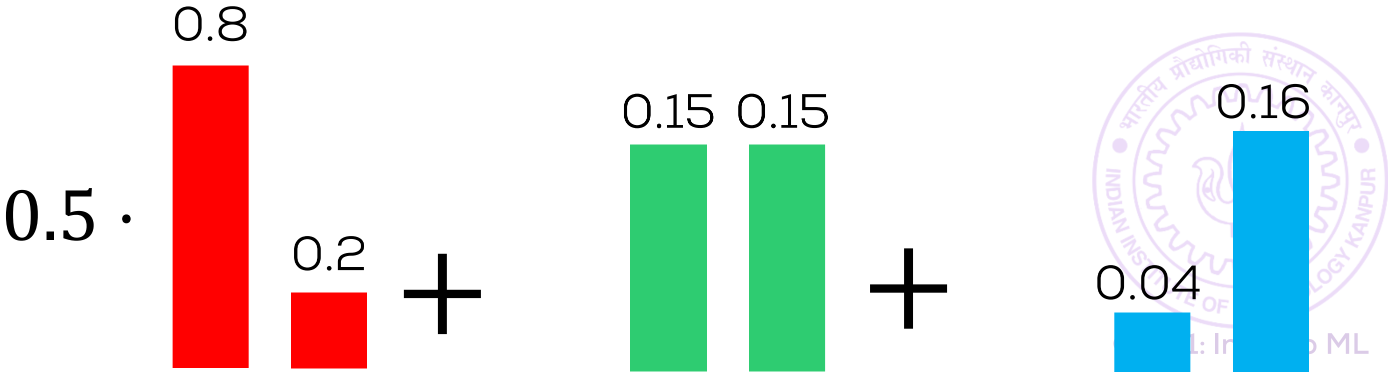


"Mixture" of 3 Bernoulli's

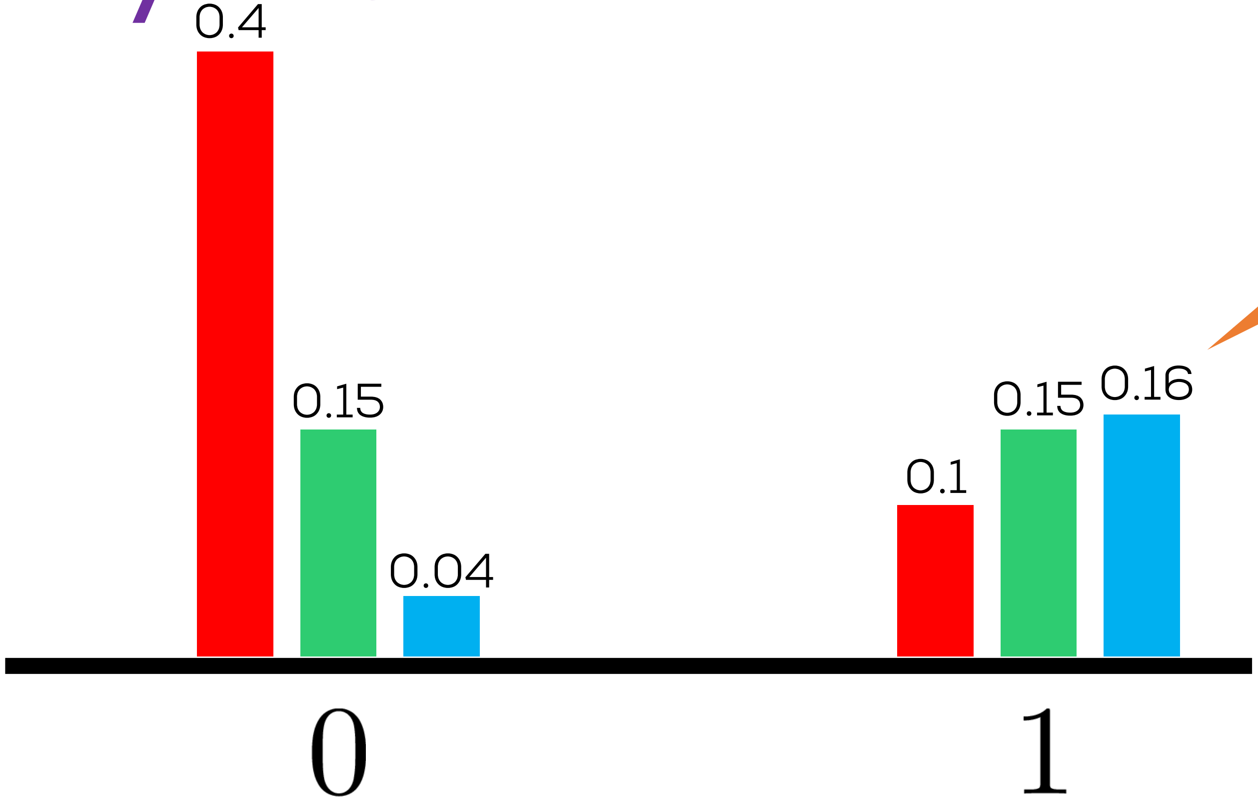
$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not



My First Generative Model

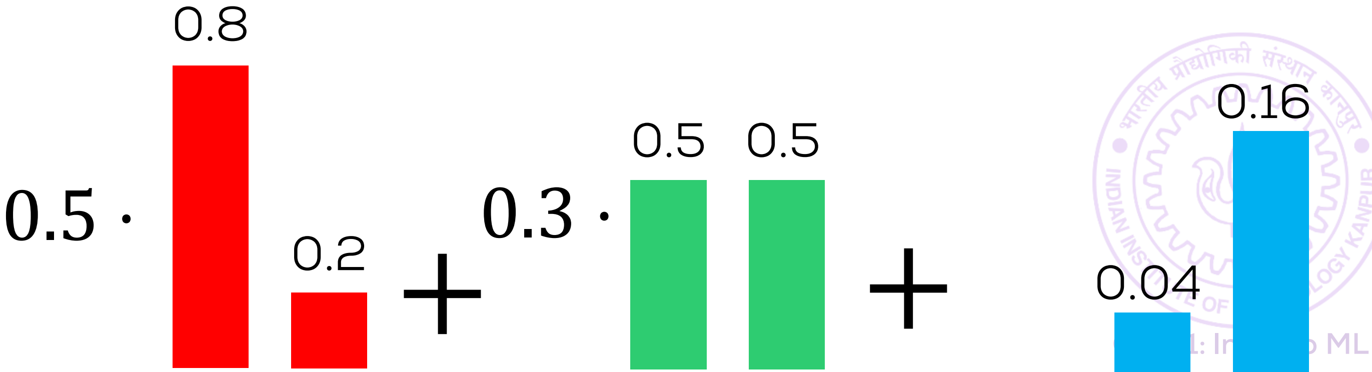


"Mixture" of 3 Bernoulli's

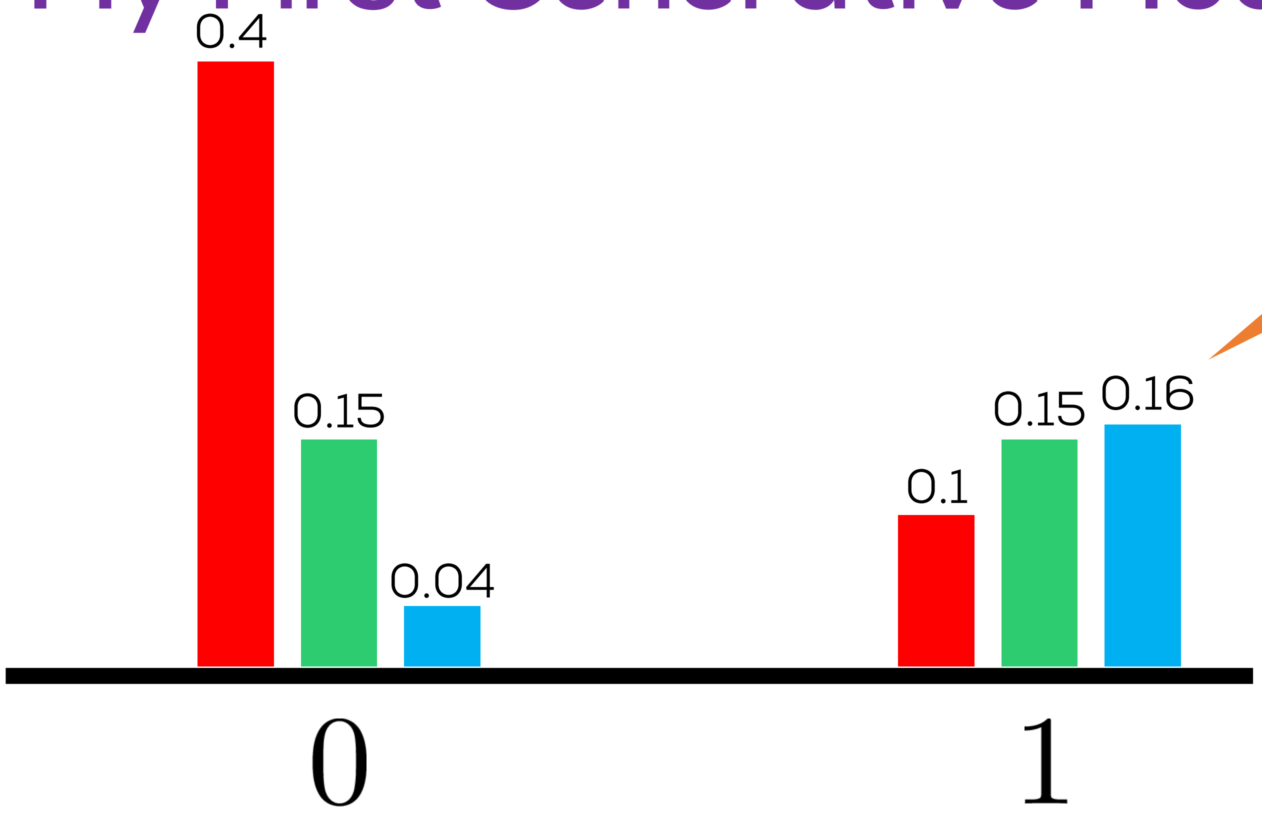
$\boldsymbol{x} \in \{0,1\}$, $y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not



My First Generative Model



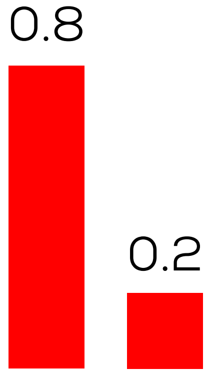
"Mixture" of 3 Bernoulli's

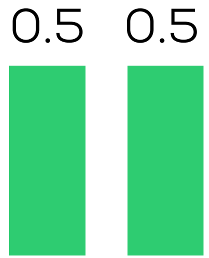
$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

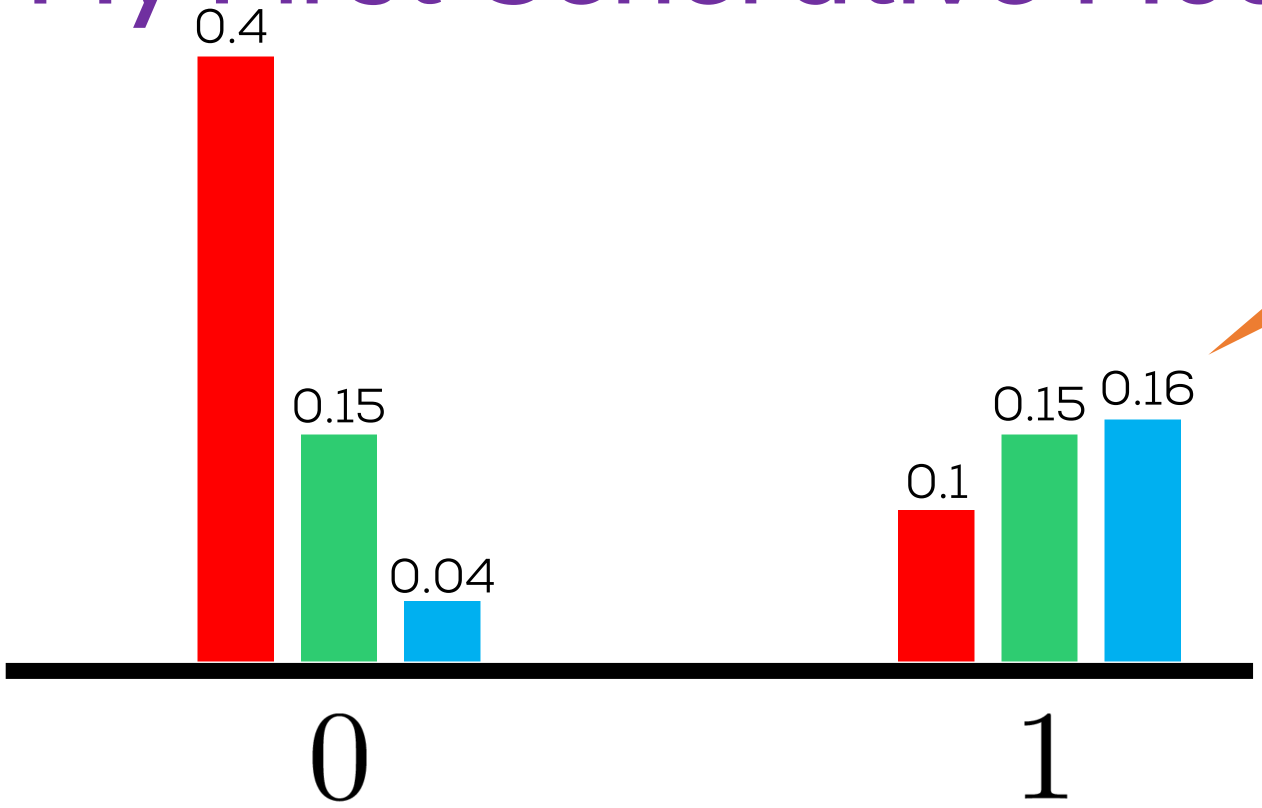
Class proportion/ weight/popularity/ imbalance

$0.5 \cdot$ 

$+ \; 0.3 \cdot$ 

$+ \; 0.2 \cdot$ 

My First Generative Model

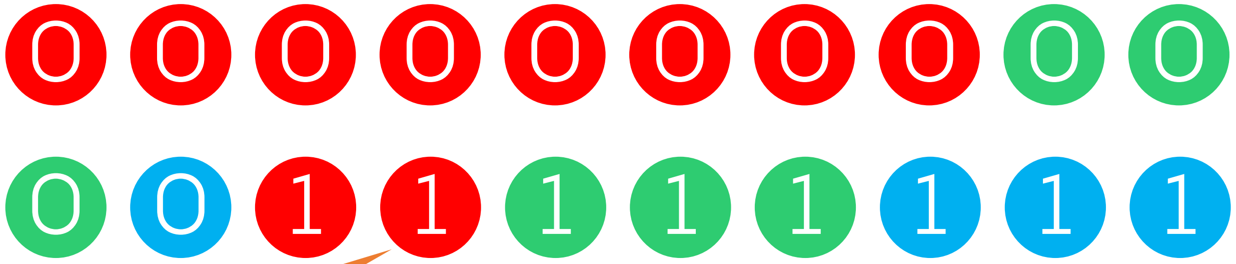


“Mixture” of 3 Bernoulli’s

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{\color{red}{1}, \color{green}{2}, \color{cyan}{3}\}$

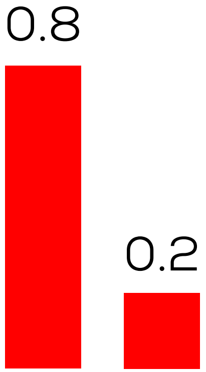
Can I learn this mixture?

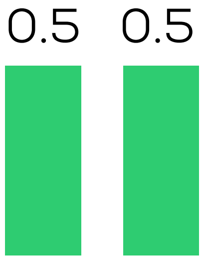
E.g. x can denote whether the word “urgent” appears in the mail or not



Training data

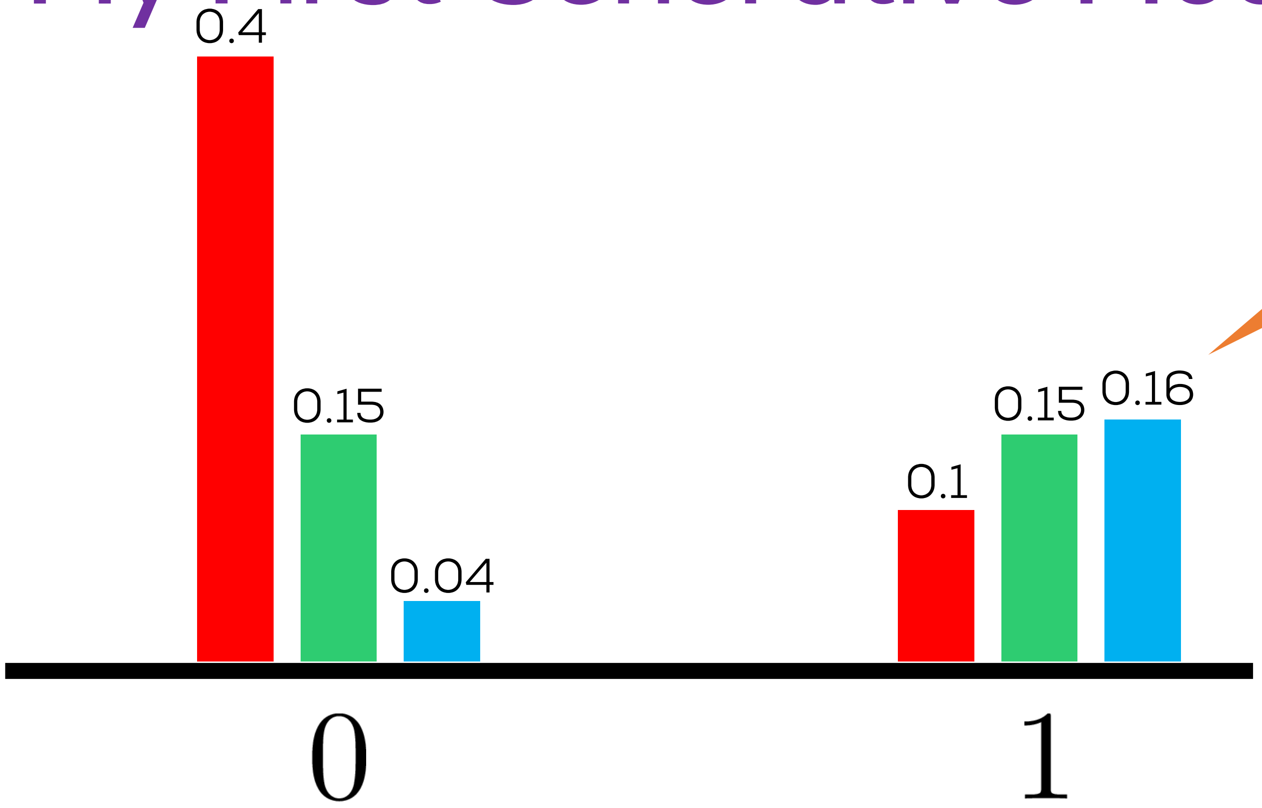
Class proportion/ weight/popularity/ imbalance

$0.5 \cdot$


$+\ 0.3 \cdot$


$+\ 0.2 \cdot$


My First Generative Model



"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

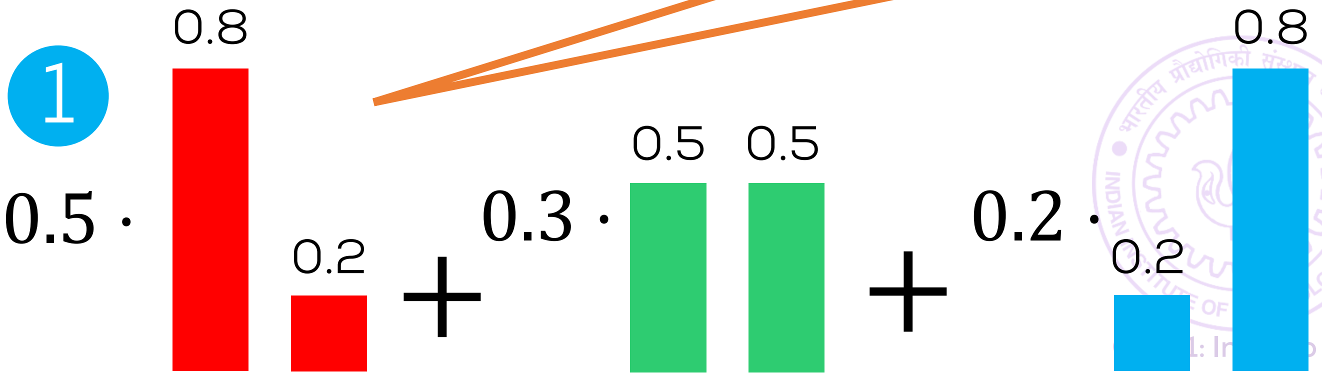
Can we estimate these parameters using training data?

0 0 0 0 0 0 0 0 0 0

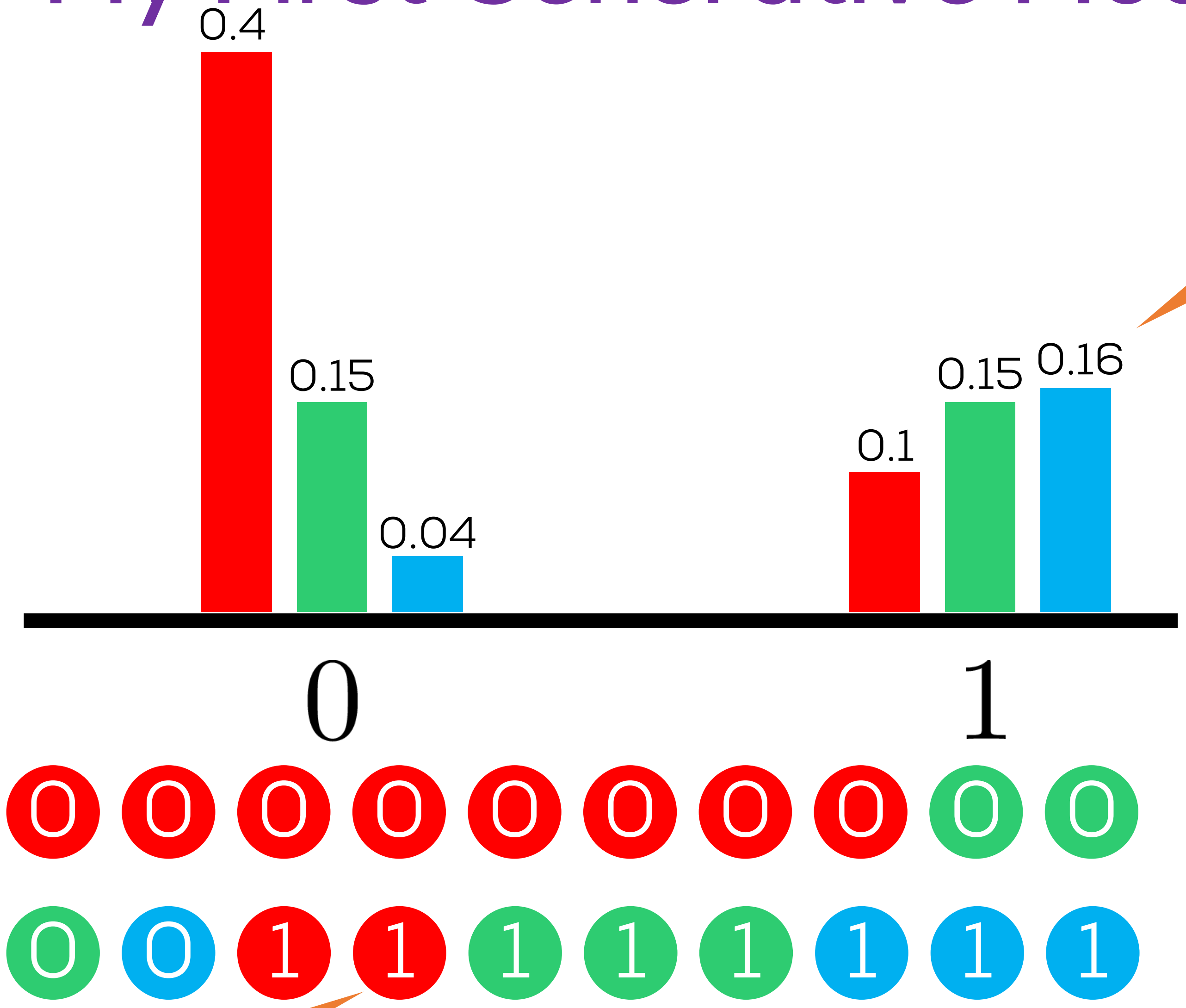
0 0 1 1 1 1 1 1 1 1

Training data

Class proportion/ weight/popularity/ imbalance



My First Generative Model



Training data

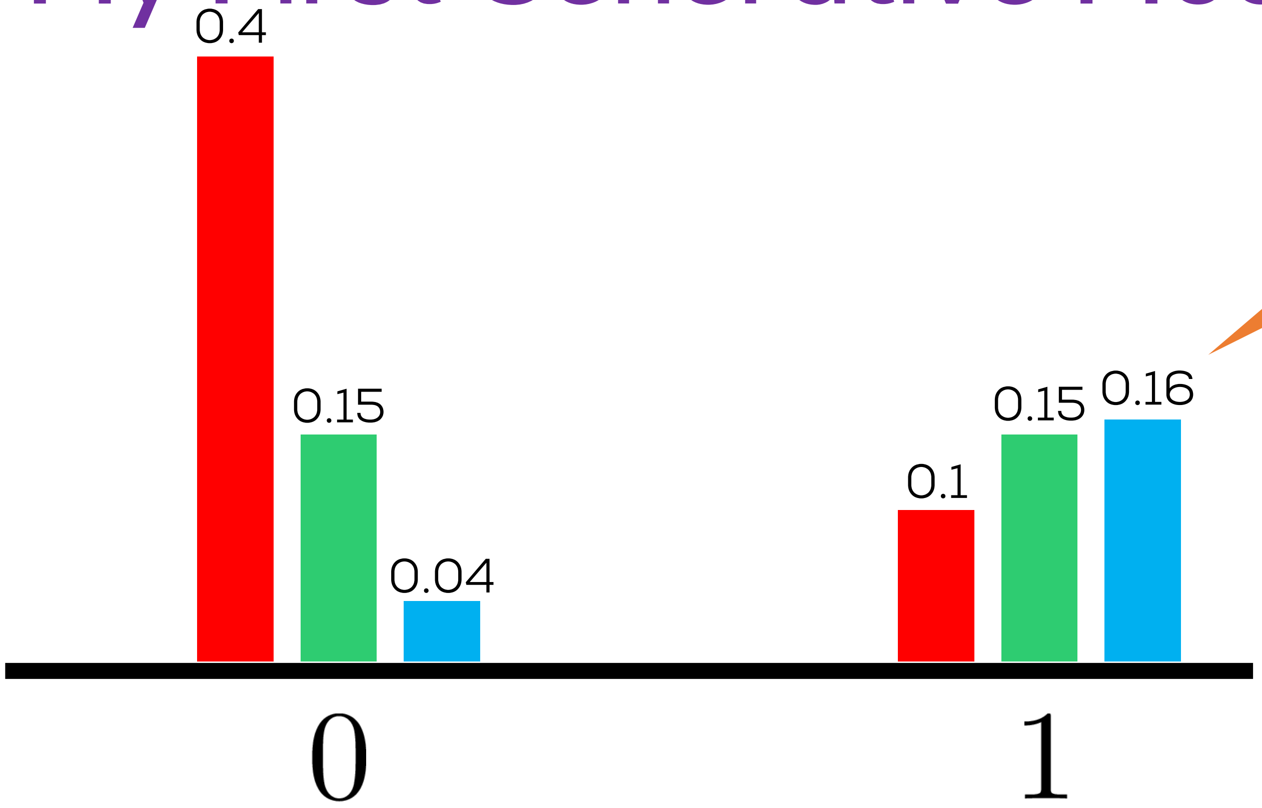
"Mixture" of 3 Bernoulli's

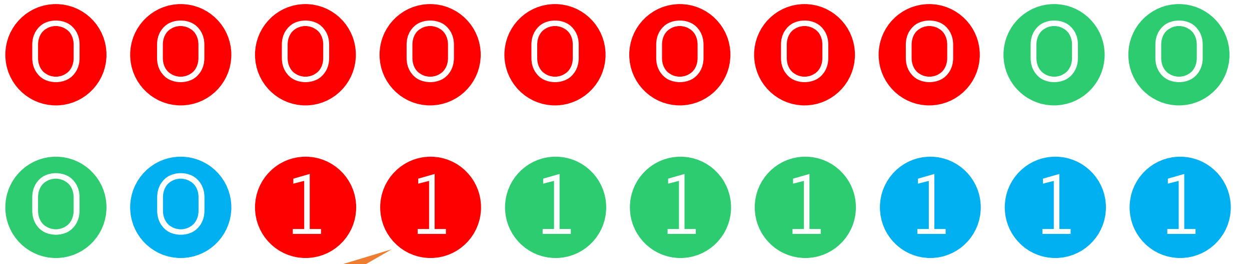
$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

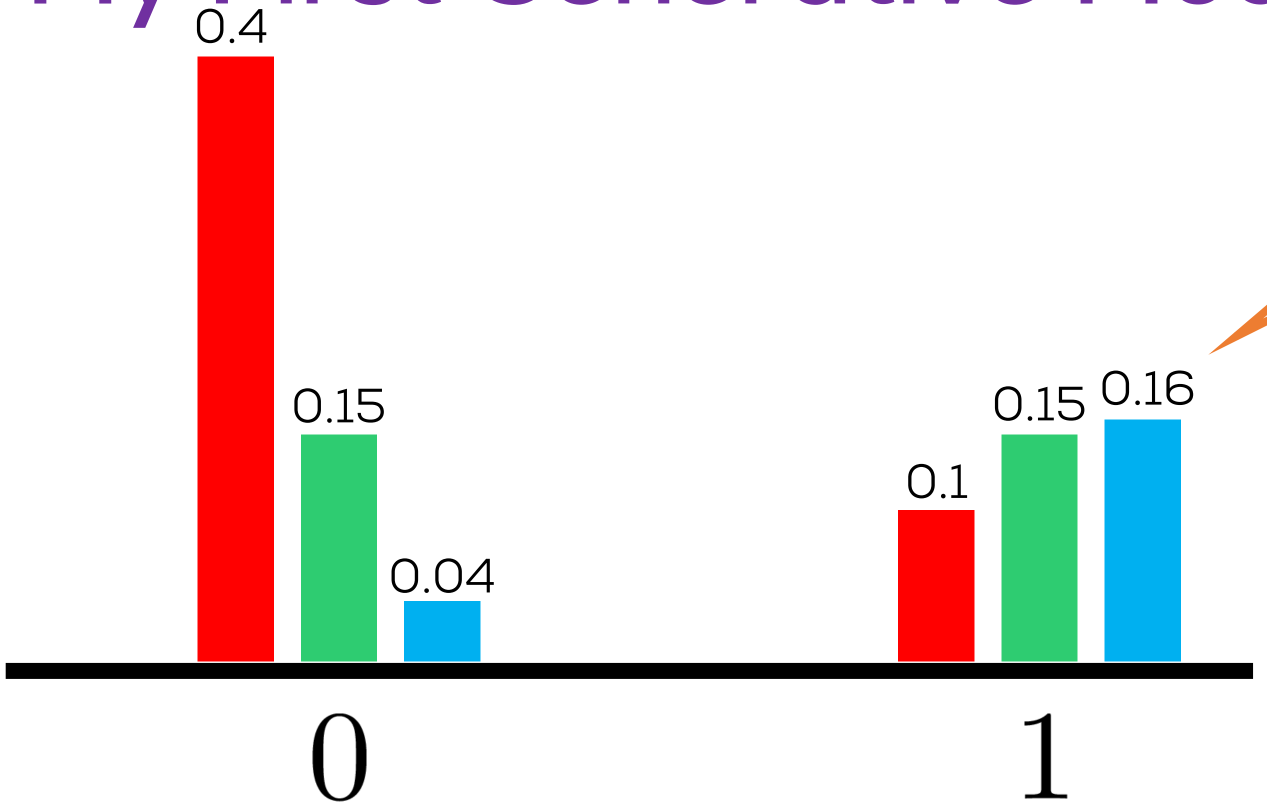
$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

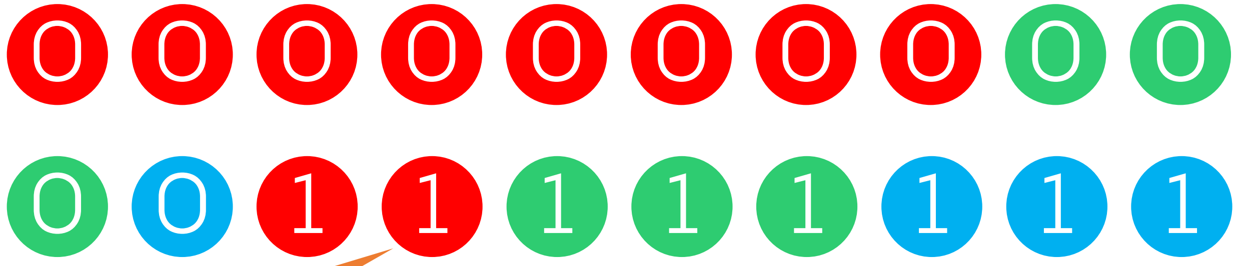
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

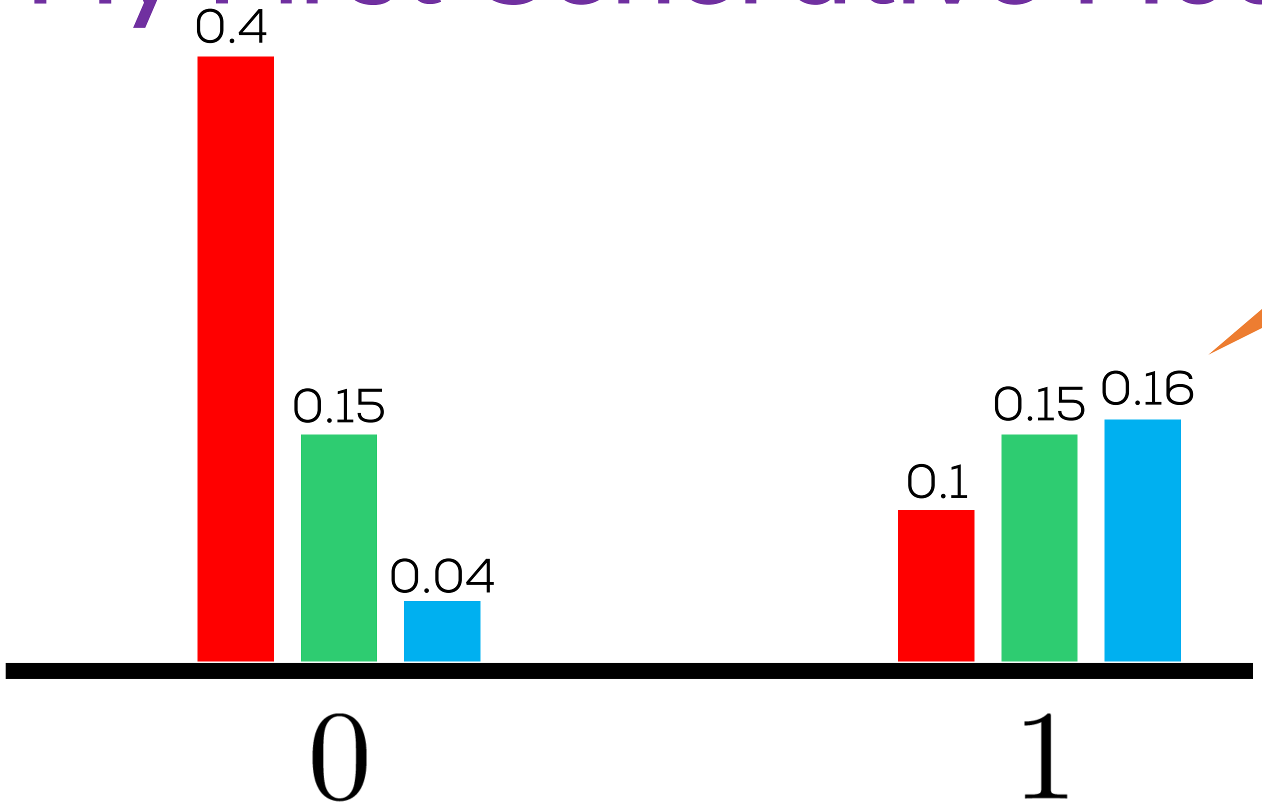
Can I learn this mixture?

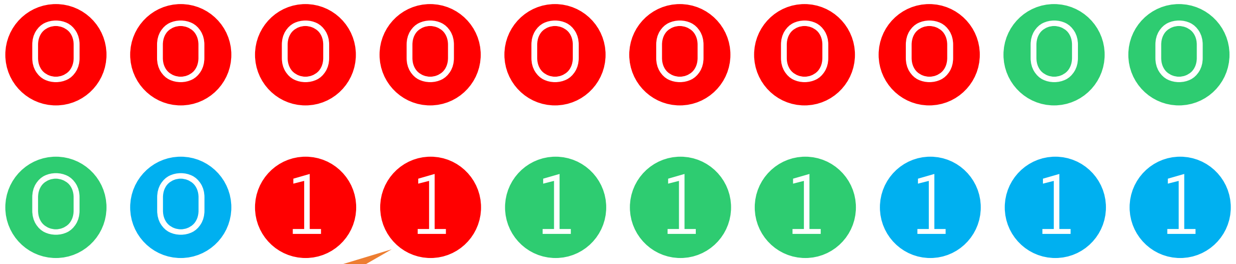
E.g. x can denote whether the word "urgent" appears in the mail or not

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

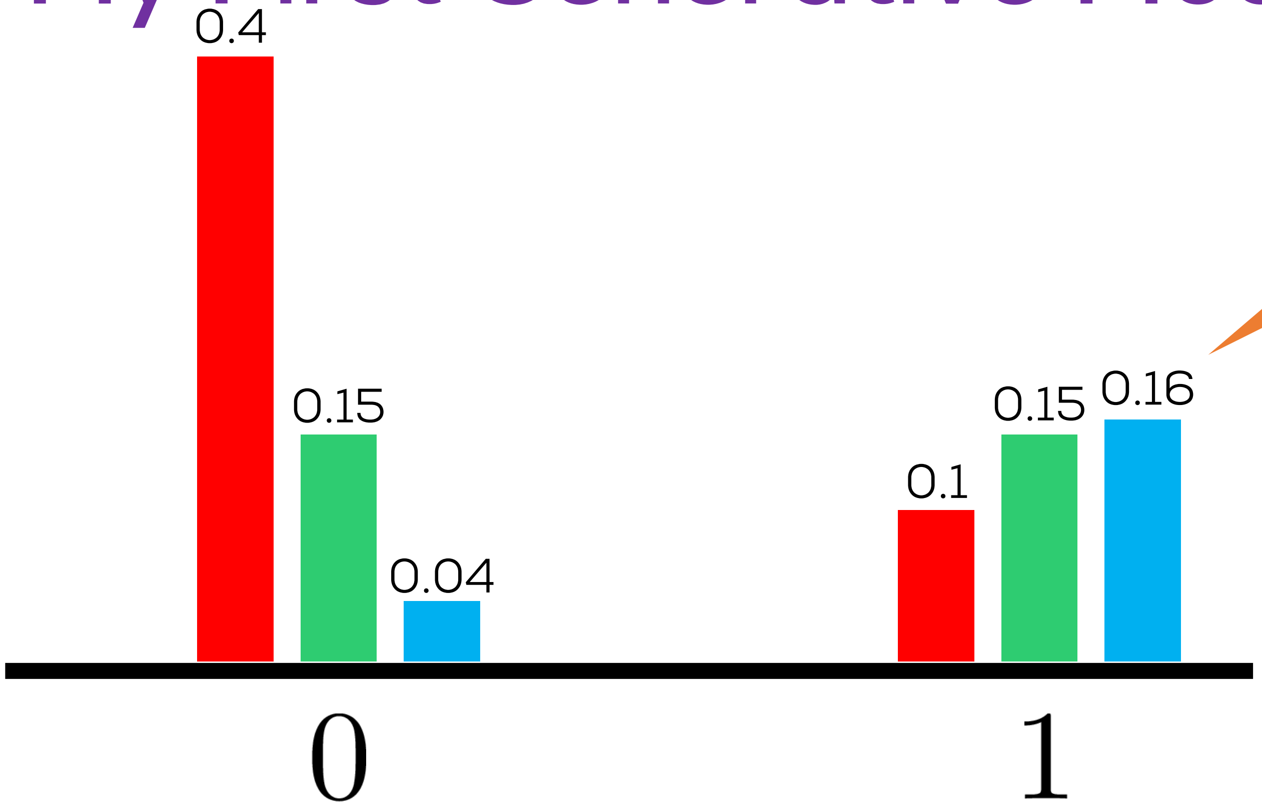
Can I learn this mixture?

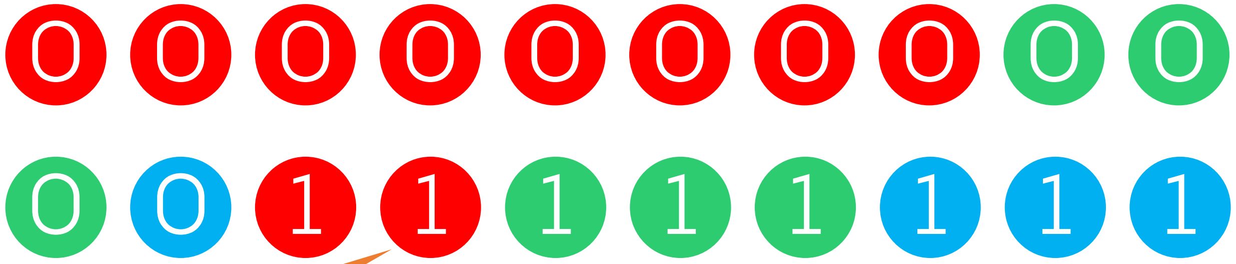
E.g. x can denote whether the word "urgent" appears in the mail or not

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

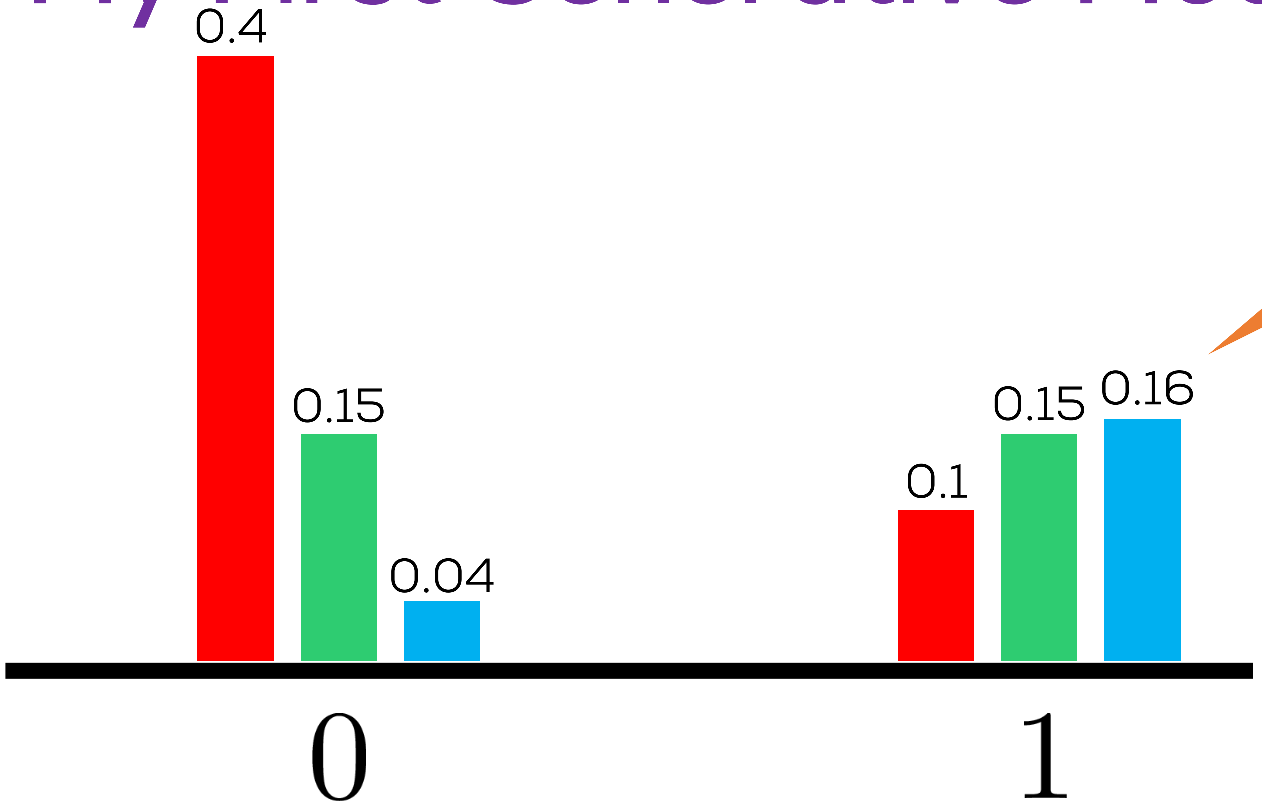
Can I learn this mixture?

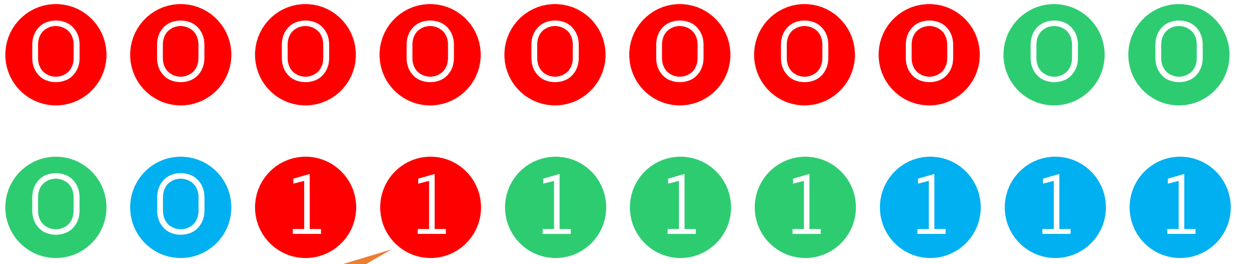
E.g. x can denote whether the word "urgent" appears in the mail or not

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$x \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

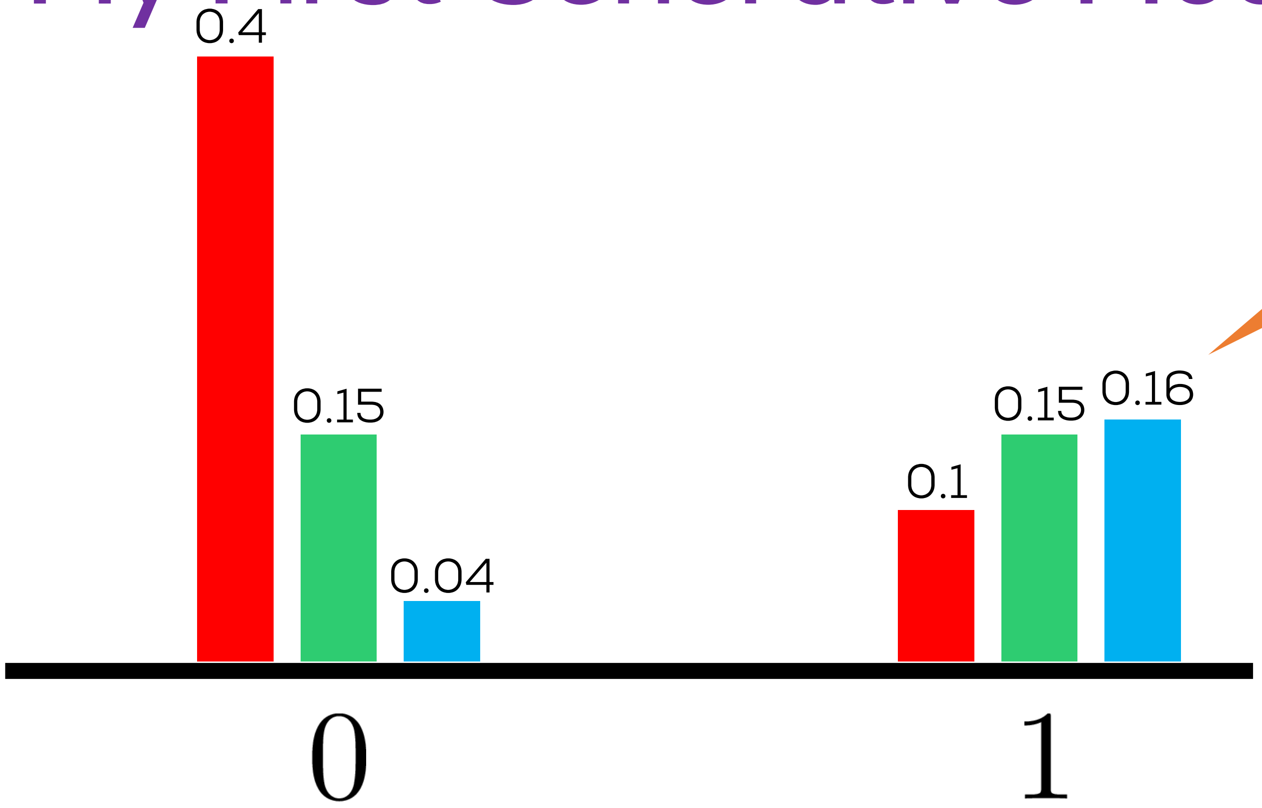
E.g. x can denote whether the word "urgent" appears in the mail or not

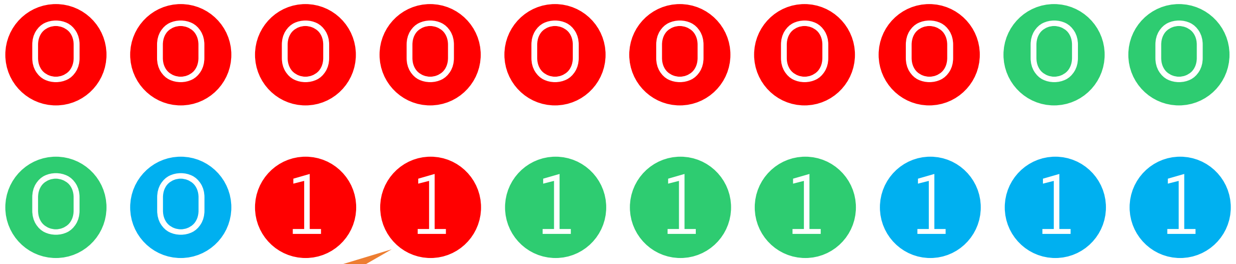
Can be shown to be the MLE!

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

My First Generative Model





Training data

“Mixture” of 3 Bernoulli’s

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word “urgent” appears in the mail or not

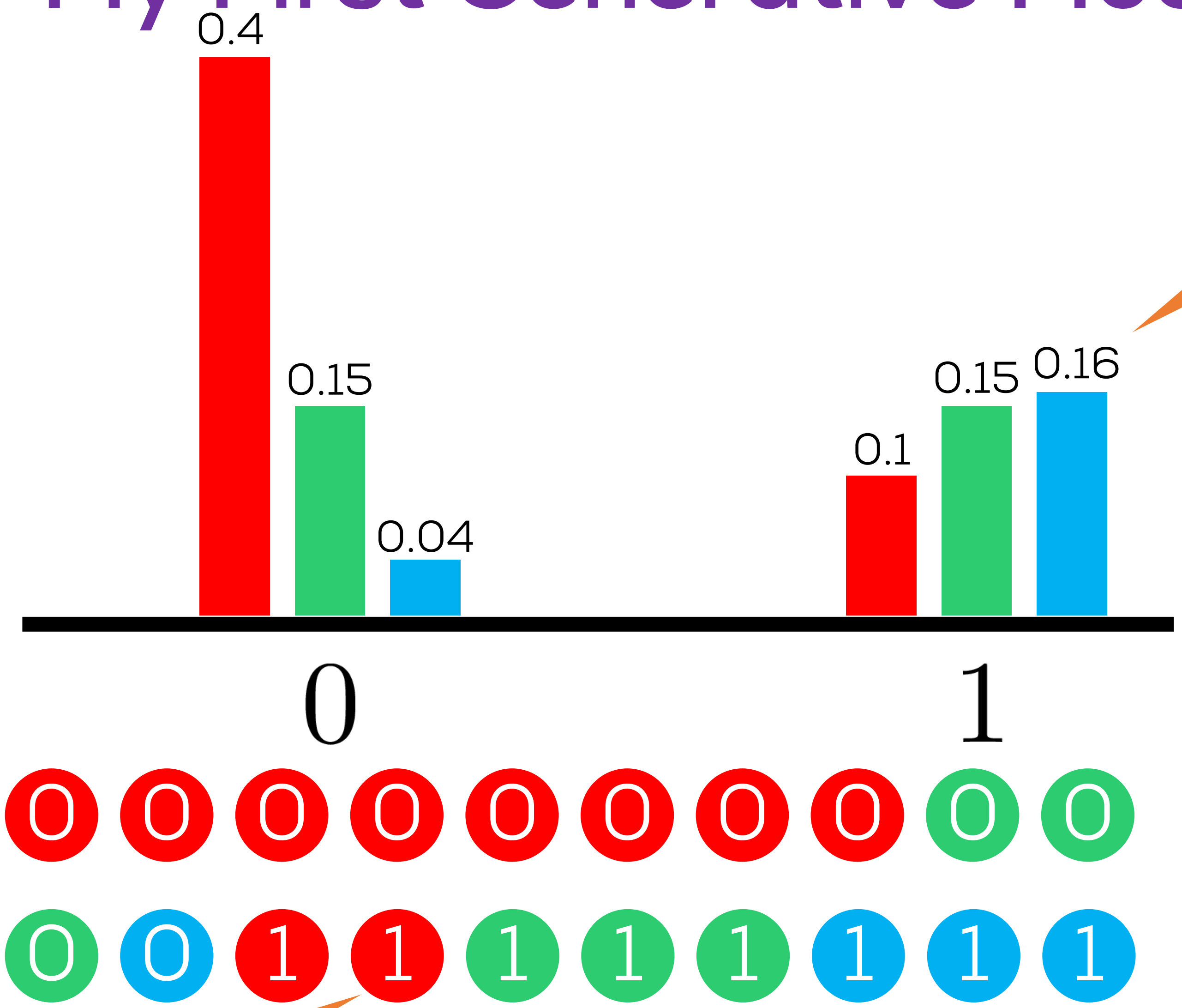
Can be shown to be the MLE!

Can do Bayesian inference too!

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

My First Generative Model



Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

Can be shown to be the MLE!

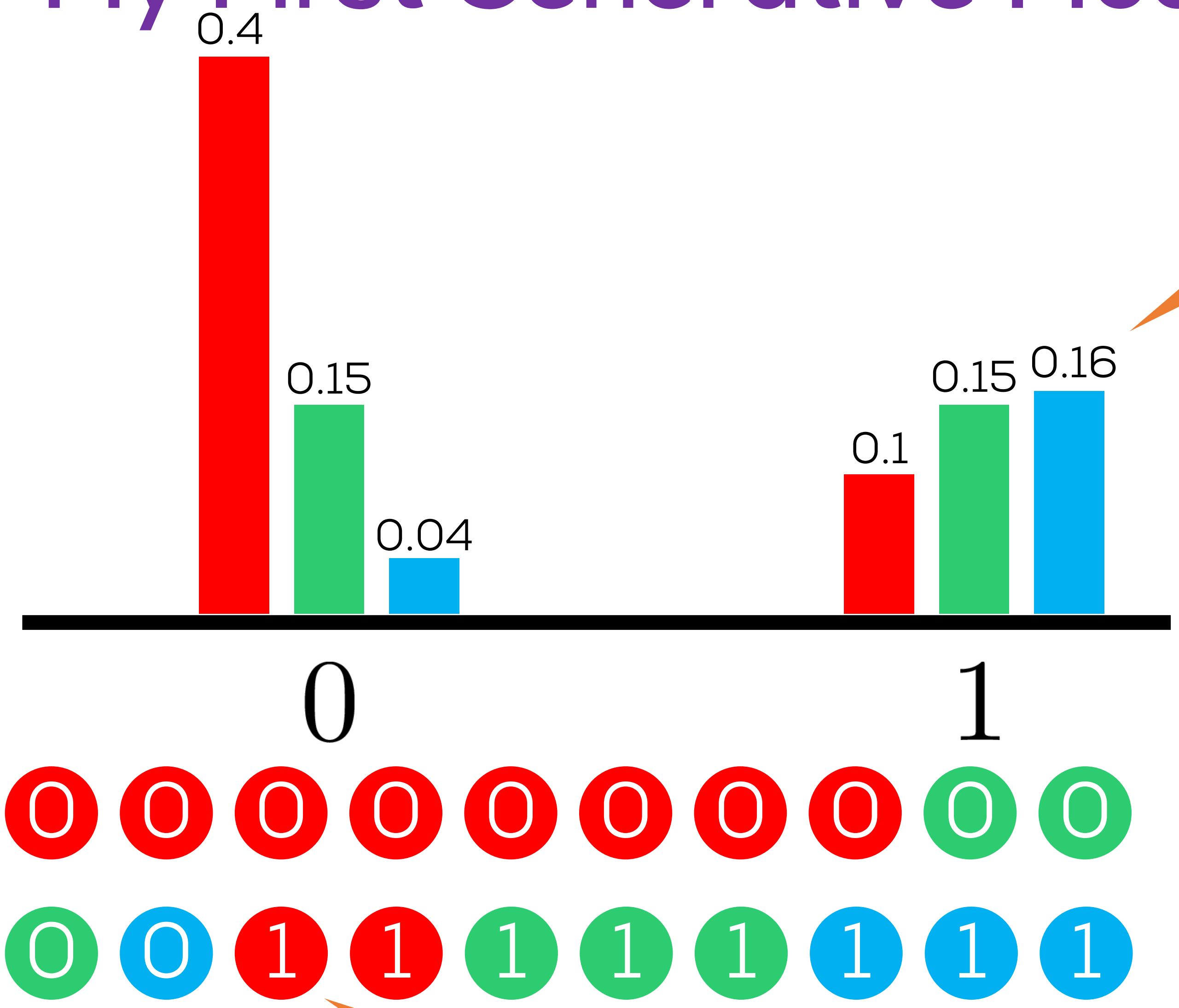
Can do Bayesian inference too!

$$\mathbb{P}[\bullet] \approx \frac{|i : y^i = \bullet|}{n}$$

Total number of data points

Exercise

My First Generative Model



"Mixture" of 3 Bernoulli's

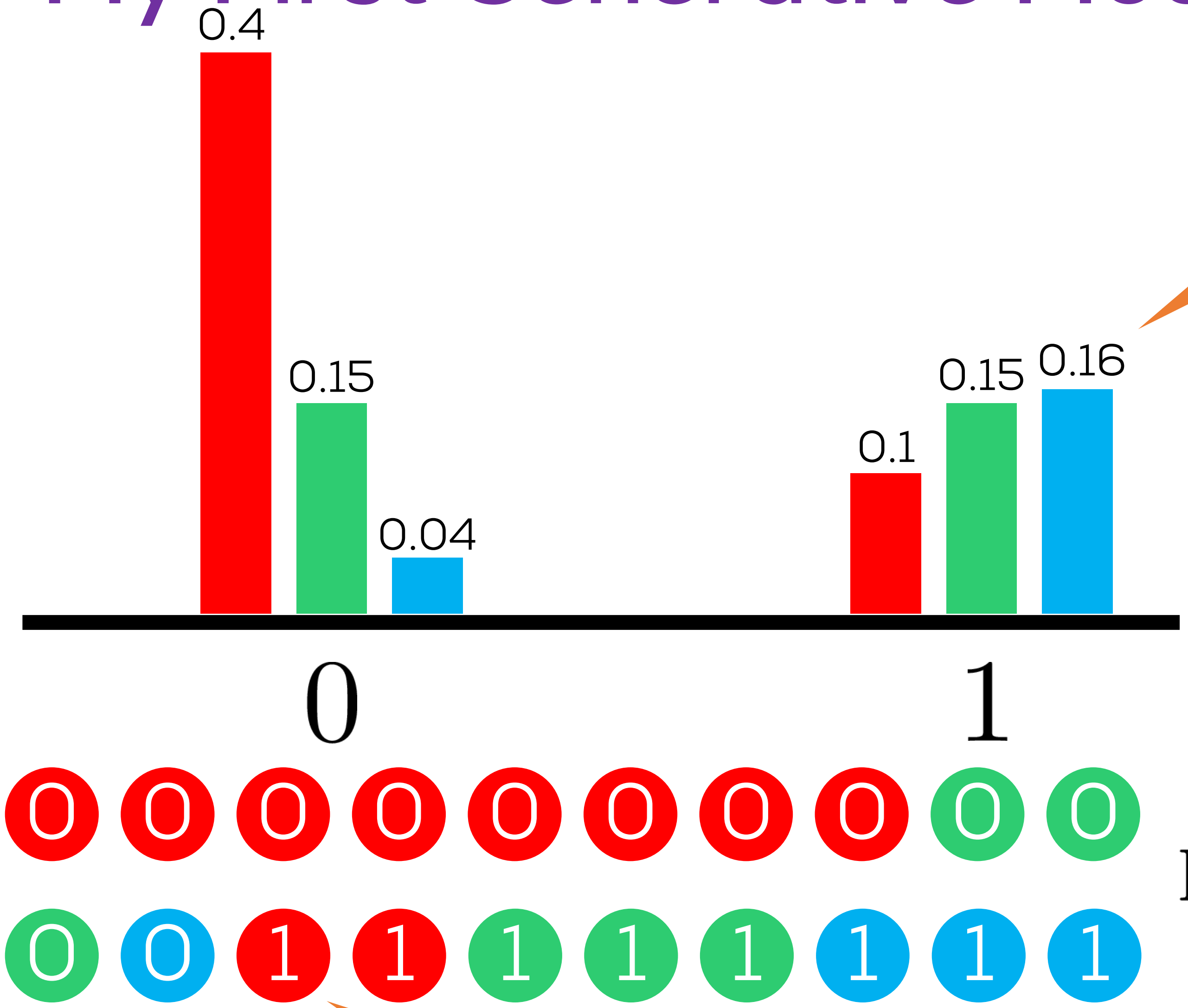
$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

Training data

My First Generative Model



"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

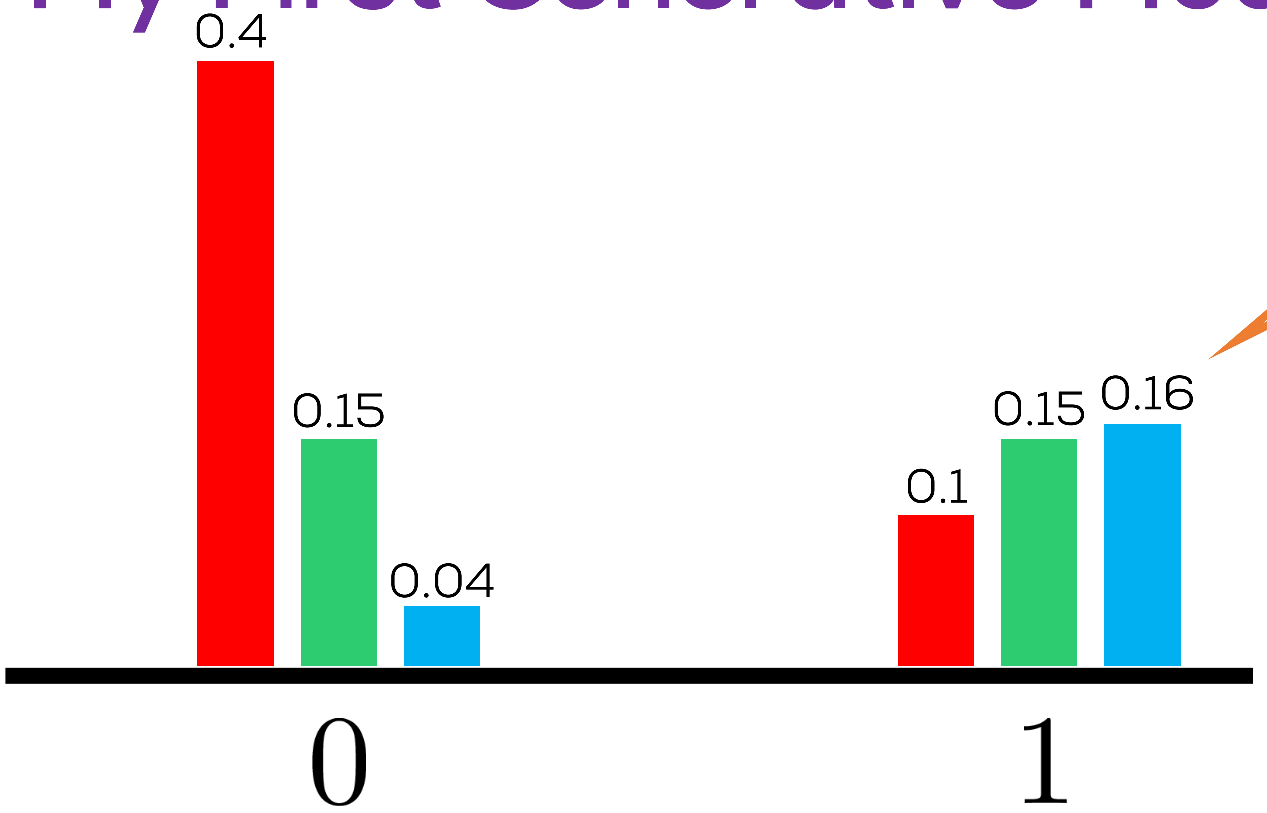
Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

$$\mathbb{P}\left[x = 0 \,|\, \bullet\right] \approx \frac{\left|i : x^i = 0 \cap y^i = \bullet\right|}{\left|i : y^i = \bullet\right|}$$

Training data

My First Generative Model

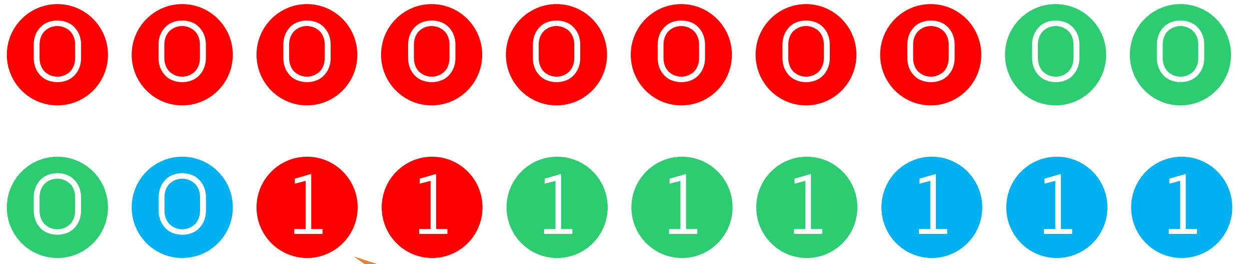


"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

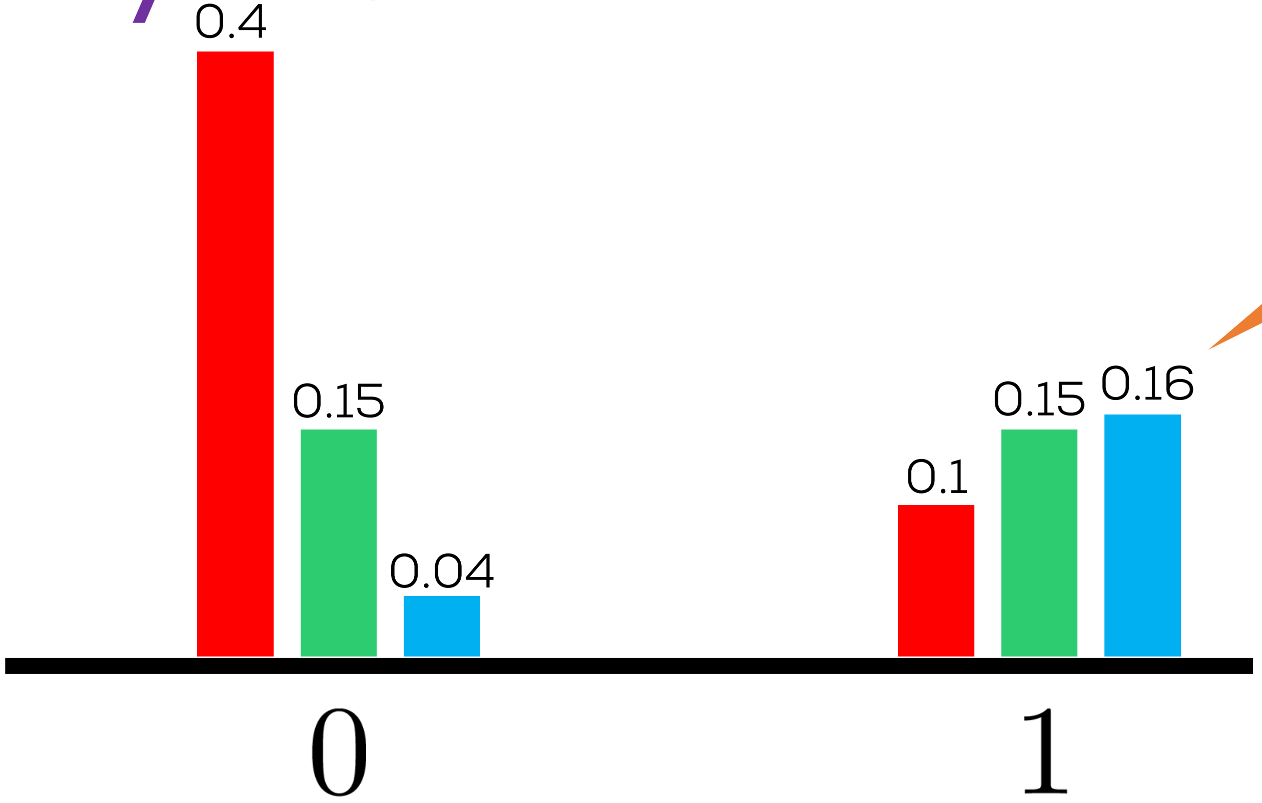


Training data

$$\mathbb{P}[x = 0 \mid \bullet] \approx \frac{|i : x^i = 0 \cap y^i = \bullet|}{|i : y^i = \bullet|}$$

Total number of data points with label $\bullet$

My First Generative Model



"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 1 1 1 1

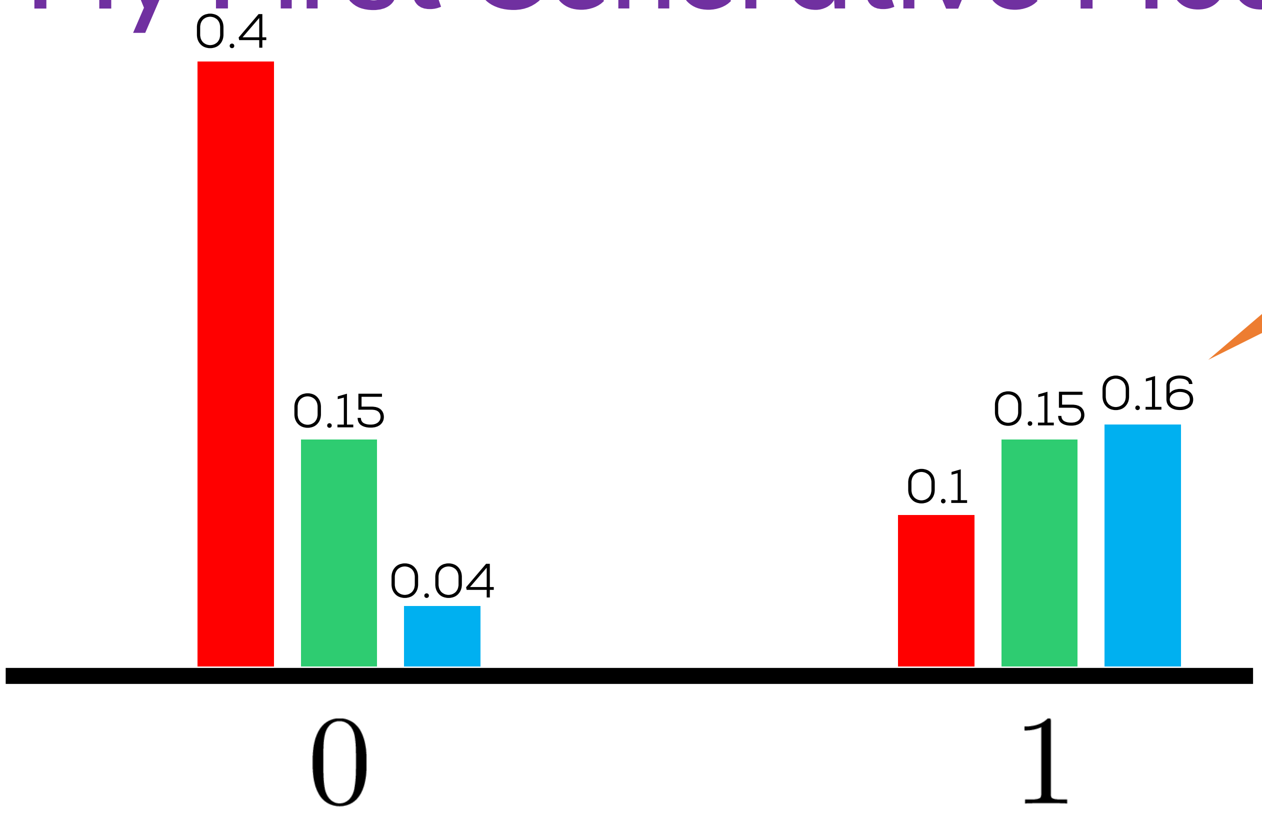
Training data

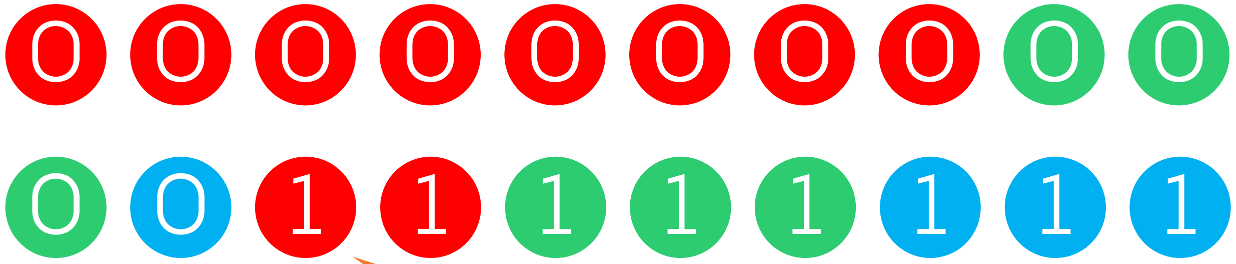
$\mathbb{P}[x = 1 | \bullet] = 1 - \mathbb{P}[x = 0 | \bullet]$

$$\mathbb{P}\left[x = 0 \,|\, \bullet\right] \approx \frac{\left|i : x^i = 0 \cap y^i = \bullet\right|}{\left|i : y^i = \bullet\right|}$$

Total number of data points with label ●

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

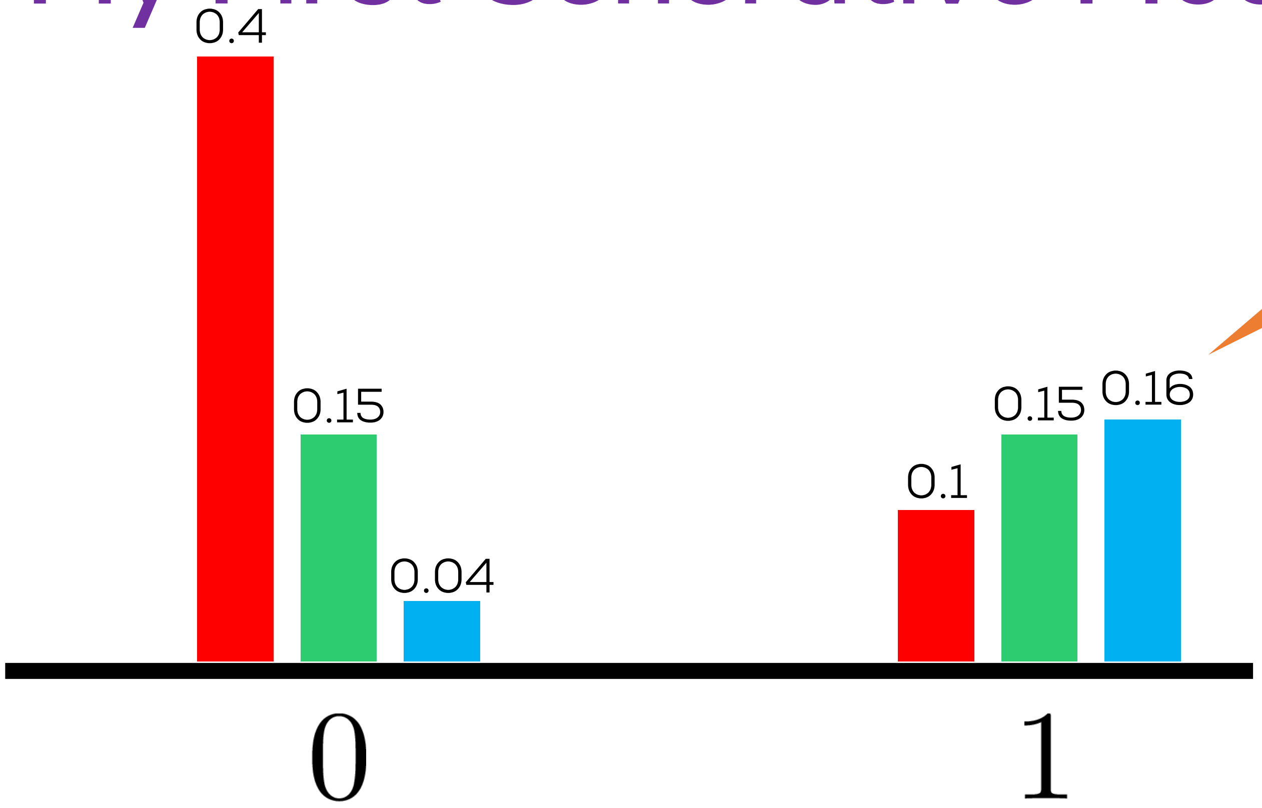
E.g. x can denote whether the word "urgent" appears in the mail or not

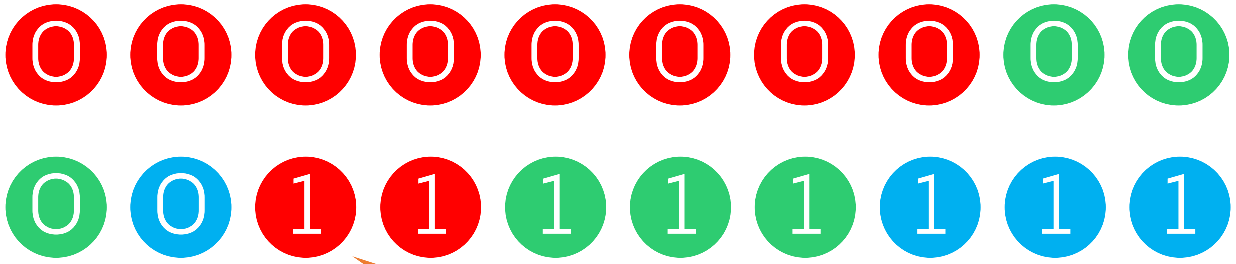
$\mathbb{P}[x = 1 | \bullet] = 1 - \mathbb{P}[x = 0 | \bullet]$

$$\mathbb{P}\left[x = 0 \,|\, \bullet \right] \approx \frac{\left| i : x^i = 0 \cap y^i = \bullet \right|}{\left| i : y^i = \bullet \right|}$$

Total number of data points with label $\bullet$

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{{\color{red}1}, {\color{green}2}, {\color{cyan}3}\}$

Can I learn this mixture?

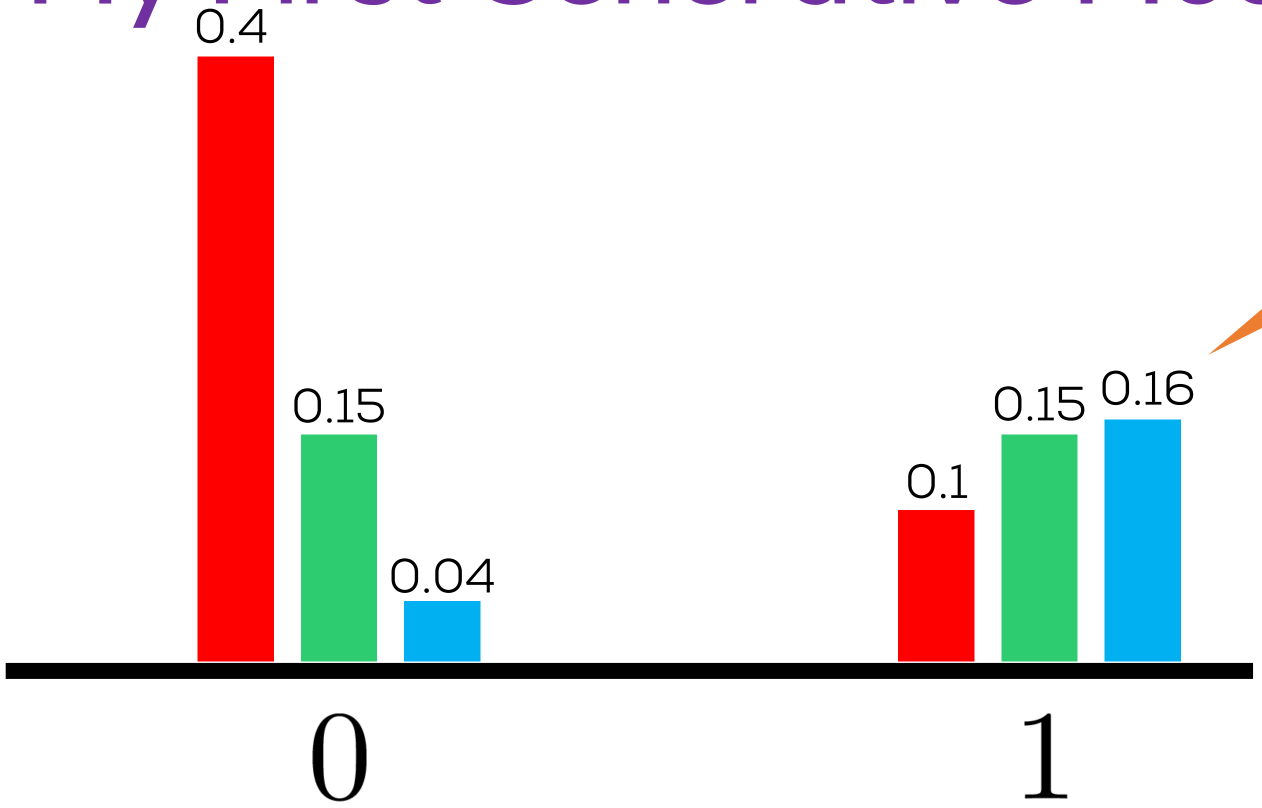
E.g. x can denote whether the word "urgent" appears in the mail or not

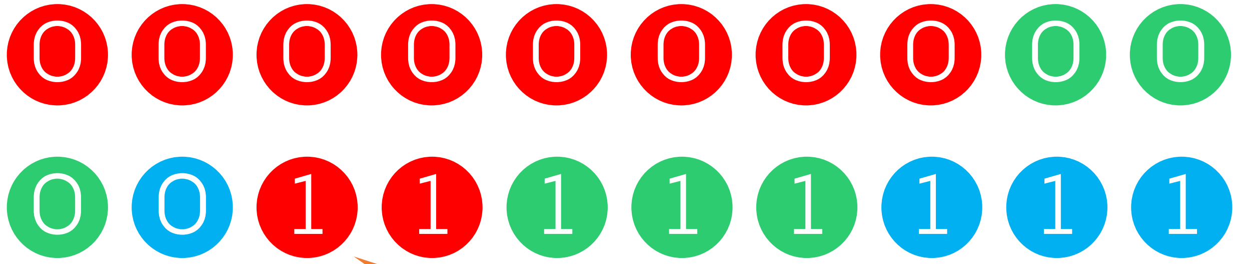
$\mathbb{P}[x = 1 | \bullet] = 1 - \mathbb{P}[x = 0 | \bullet]$

$$\mathbb{P}\left[x = 0 \,|\, \bullet\right] \approx \frac{\left|i : x^i = 0 \cap y^i = \bullet\right|}{\left|i : y^i = \bullet\right|}$$

Total number of data points with label $\bullet$

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$x \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

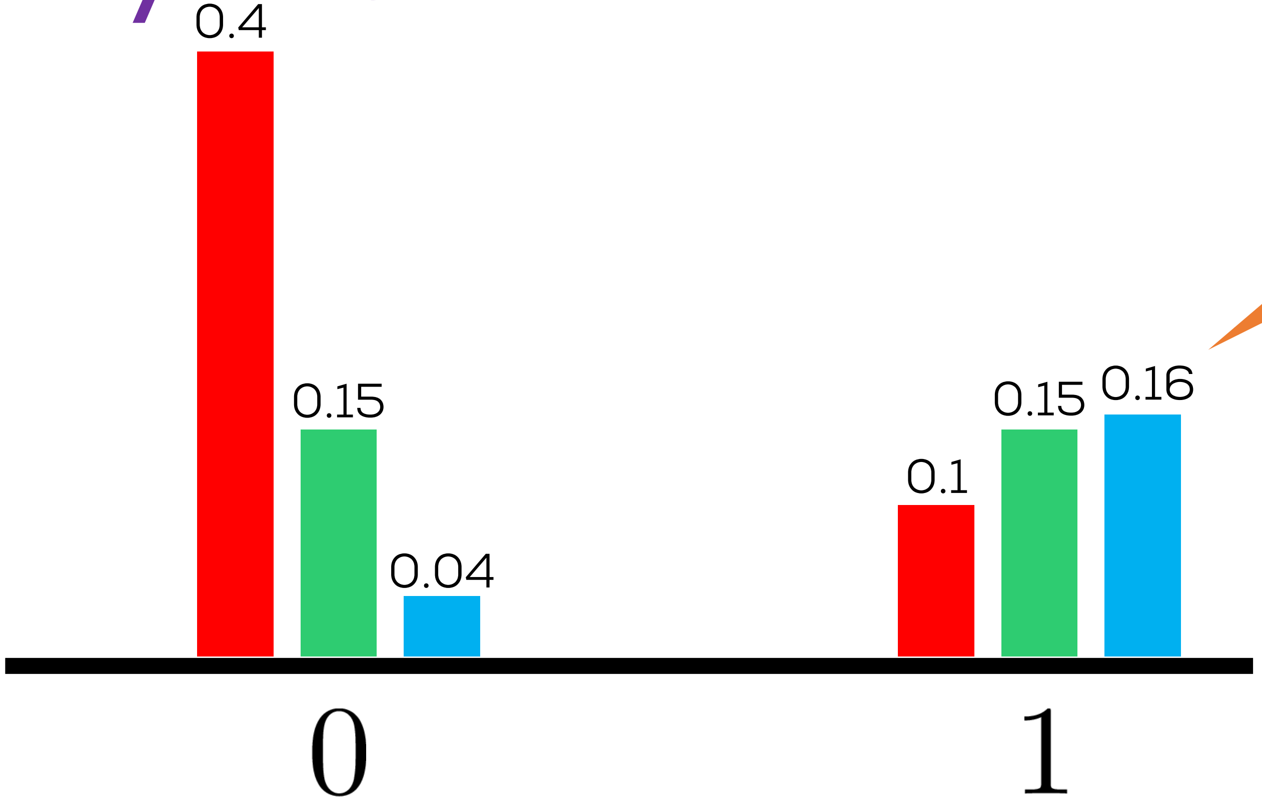
$\mathbb{P}[x = 1 | ●] = 1 - \mathbb{P}[x = 0 | ●]$

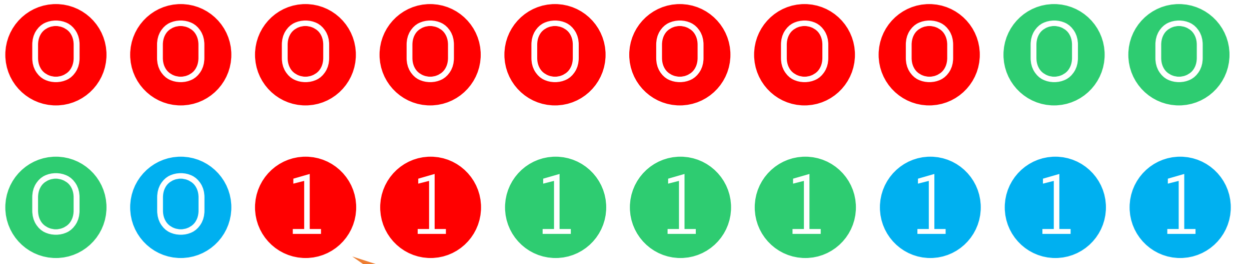
Can be shown to be the MLE!

$$\mathbb{P}\left[x = 0 \,|\, ● \right] \approx \frac{\left| i : x^i = 0 \cap y^i = ● \right|}{\left| i : y^i = ● \right|}$$

Total number of data points with label ●

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$y \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

$\mathbb{P}[x = 1 | \bullet] = 1 - \mathbb{P}[x = 0 | \bullet]$

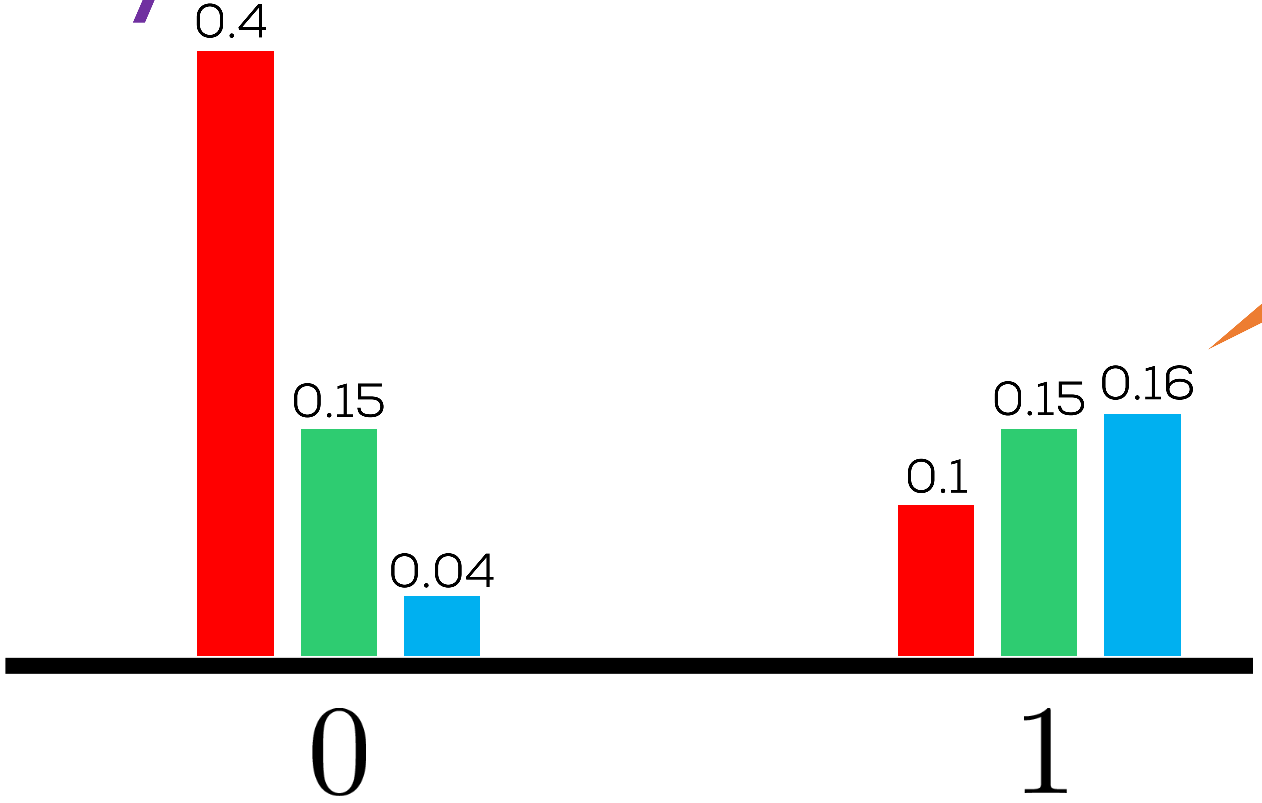
Can be shown to be the MLE!

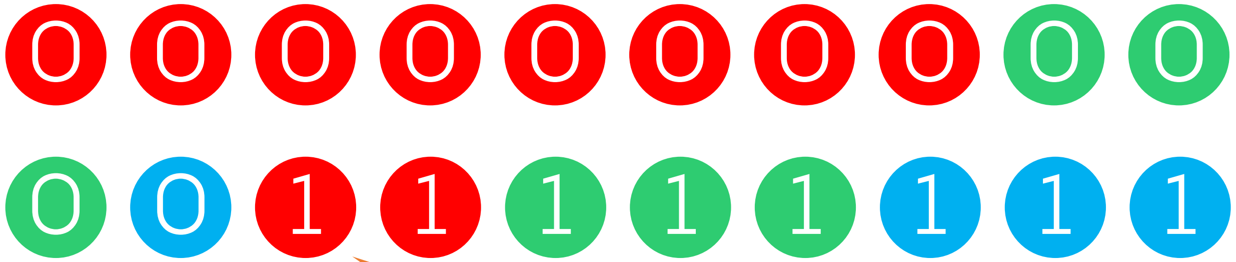
$$\mathbb{P}[x = 0 \,|\, \bullet] \approx \frac{|i : x^i = 0 \cap y^i = \bullet|}{|i : y^i = \bullet|}$$

Total number of data points with label $\bullet$

Exercise

My First Generative Model





Training data

"Mixture" of 3 Bernoulli's

$\boldsymbol{x} \in \{0,1\}$,
$\mathrm{y} \in \{1, 2, 3\}$

Can I learn this mixture?

E.g. x can denote whether the word "urgent" appears in the mail or not

$\mathbb{P}[x = 1 \mid$ ● $] = 1 - \mathbb{P}[x = 0 \mid$ ● $]$

Can be shown to be the MLE!

$x^i = 0 \cap y^i =$ ●

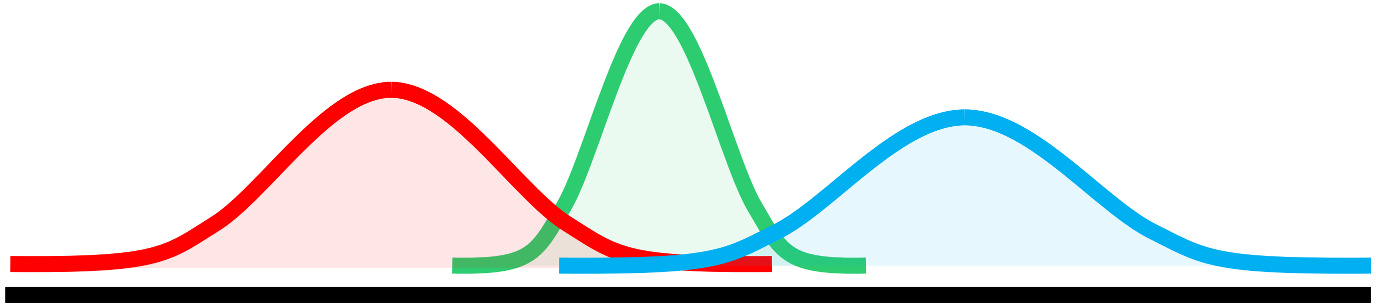
$: y^i =$ ●

Generative ≠ Bayesian

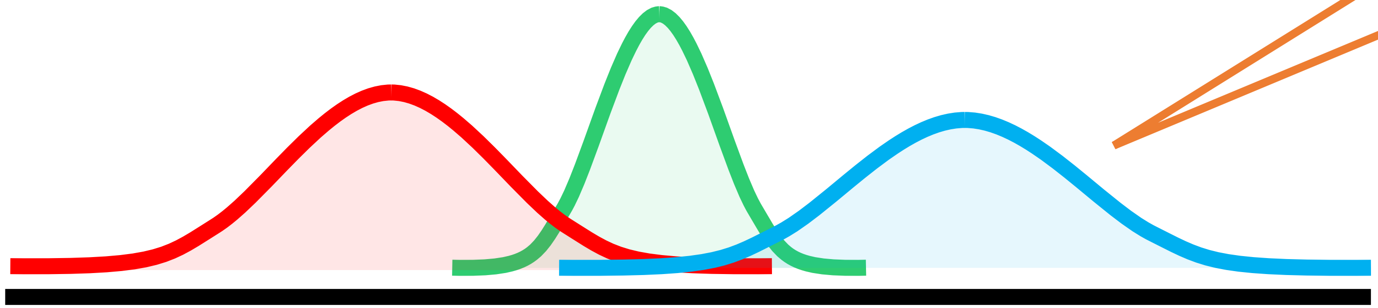
Total number of data points with label ●

Exercise

My First Generative Model

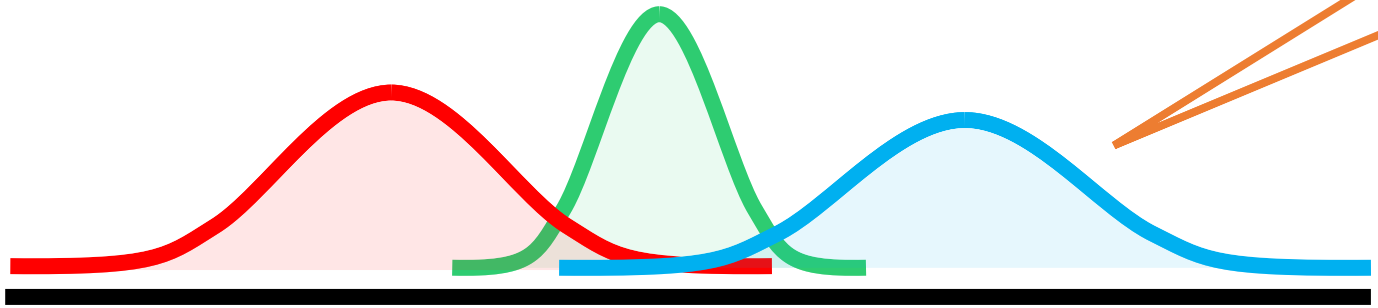


My First Generative Model



"Mixture" of
3 Gaussians

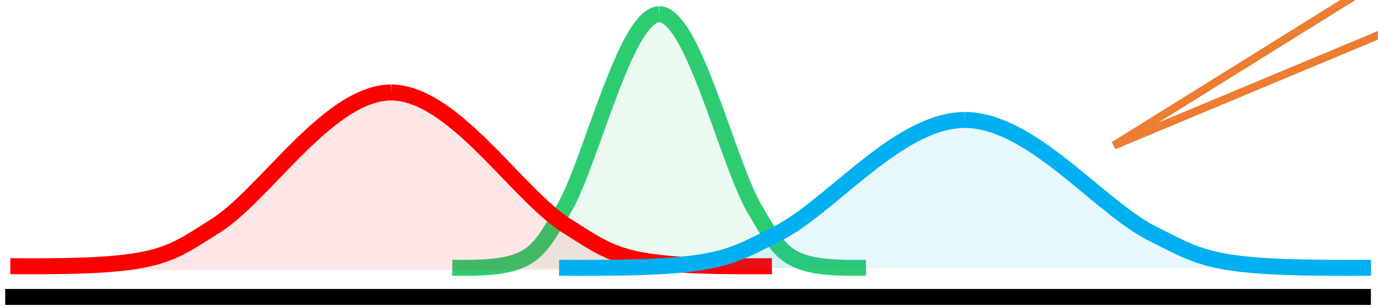
My First Generative Model



"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

My First Generative Model

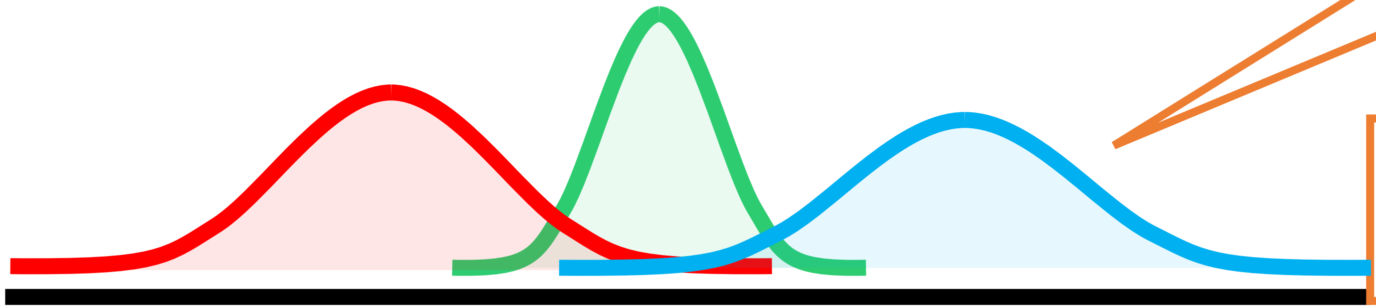


"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

My First Generative Model



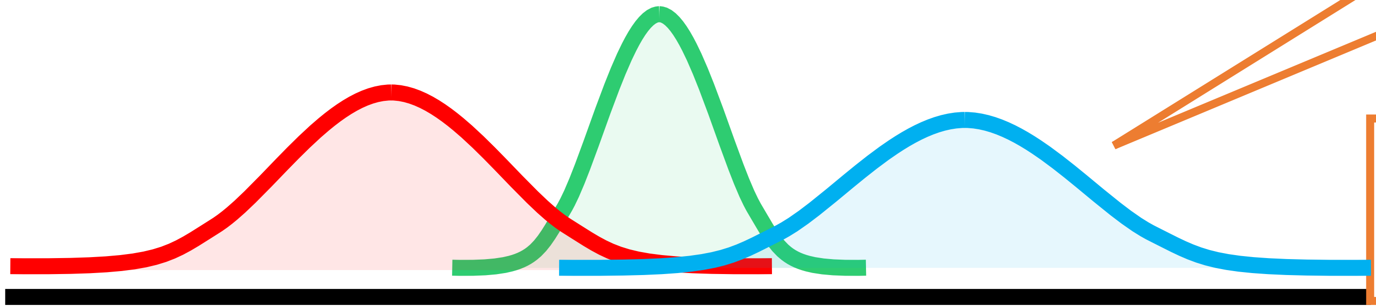
"Mixture" of 3 Gaussians

We know/assume they are Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$\mathrm{y} \in \{1, 2, 3\}$

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

My First Generative Model



"Mixture" of 3 Gaussians

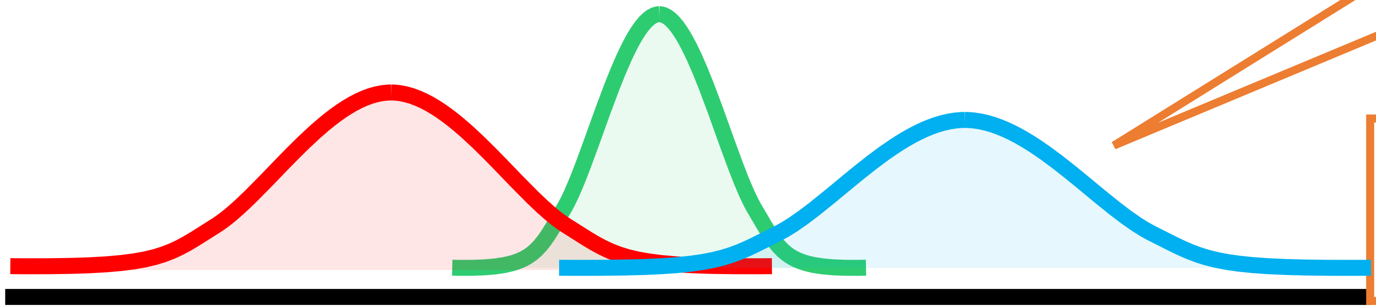
$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

We know/assume they are Gaussians

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

What we don't know/assume is what is their mean etc

My First Generative Model



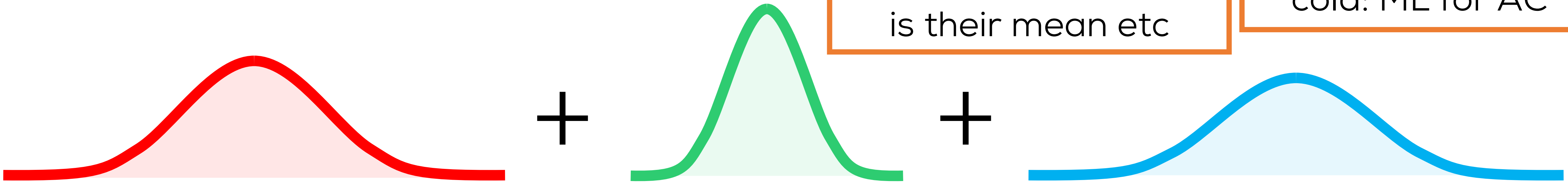
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

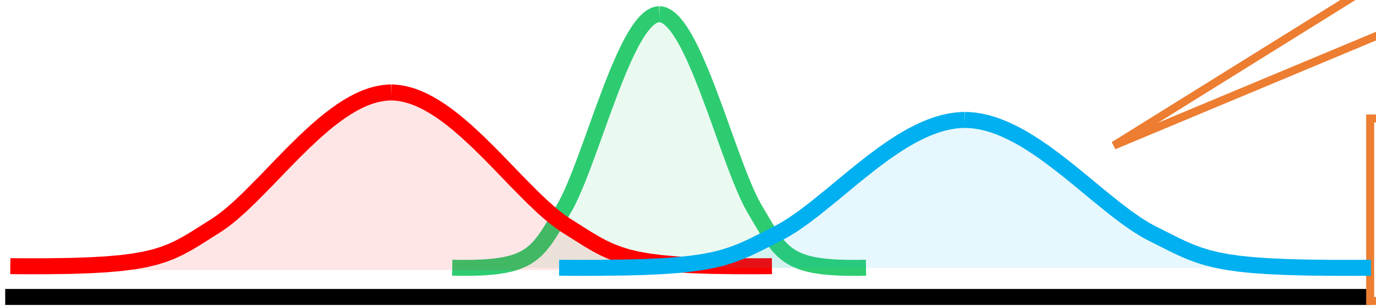
We know/assume they are Gaussians

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

What we don't know/assume is what is their mean etc



My First Generative Model



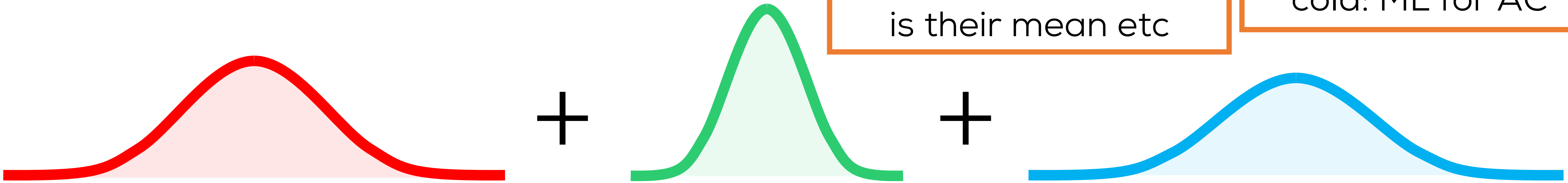
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$\mathrm{y} \in \{\color{red}{1}, \color{green}{2}, \color{cyan}{3}\}$

We know/assume they are Gaussians

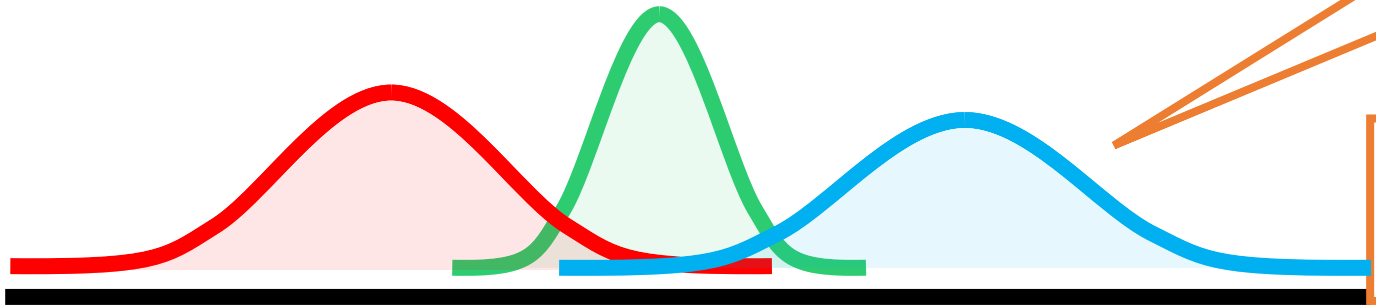
E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

What we don't know/assume is what is their mean etc



Can estimate the class proportions, and means and variances of these 1D Gaussians using training data!

My First Generative Model



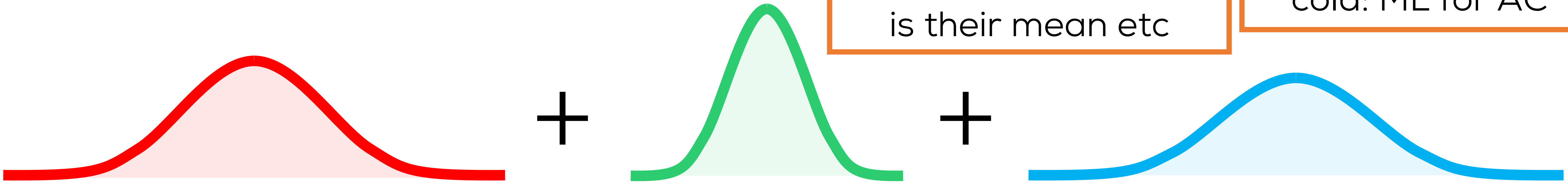
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

We know/assume they are Gaussians

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

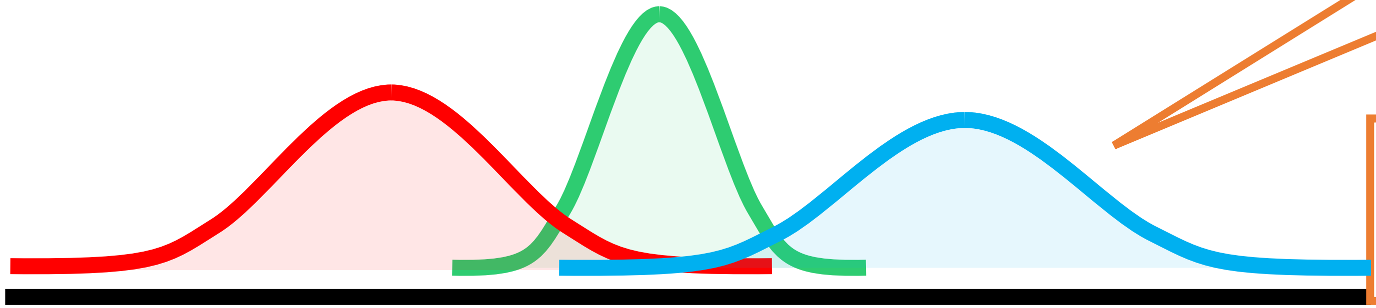
What we don't know/assume is what is their mean etc



Can estimate the class proportions, and means and variances of these 1D Gaussians using training data!

Read [DAU] Sections 9.1-9.5

My First Generative Model



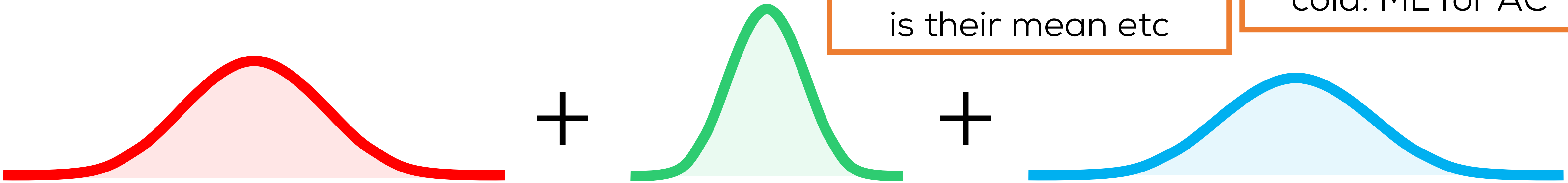
“Mixture” of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$y \in \{1, 2, 3\}$

We know/assume they are Gaussians

E.g. x could be room temperature, y could be comfort level for user, too hot, just right, too cold: ML for AC

What we don’t know/assume is what is their mean etc

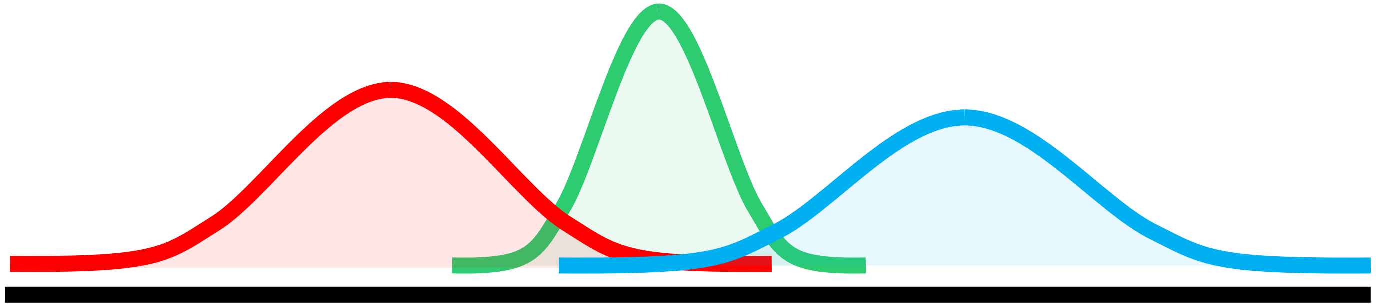


Can estimate the class proportions, and means and variances of these 1D Gaussians using training data!

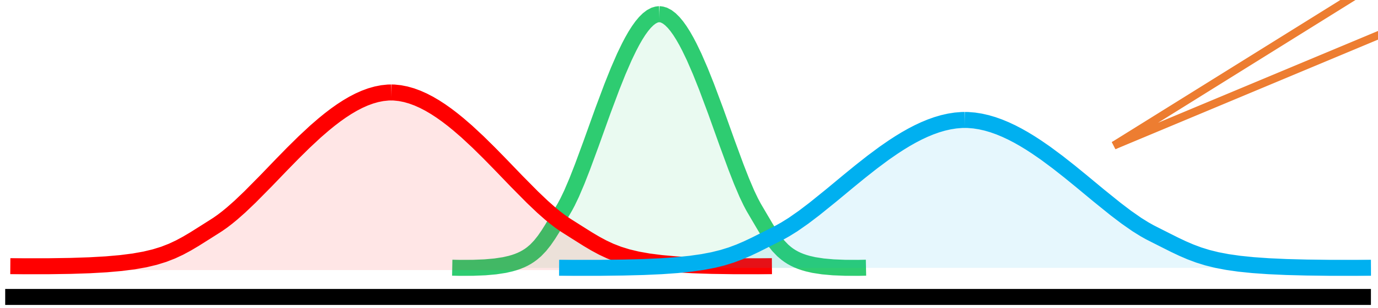
Read [DAU] Sections 9.1-9.5

Exercise

The Naïve Bayes model

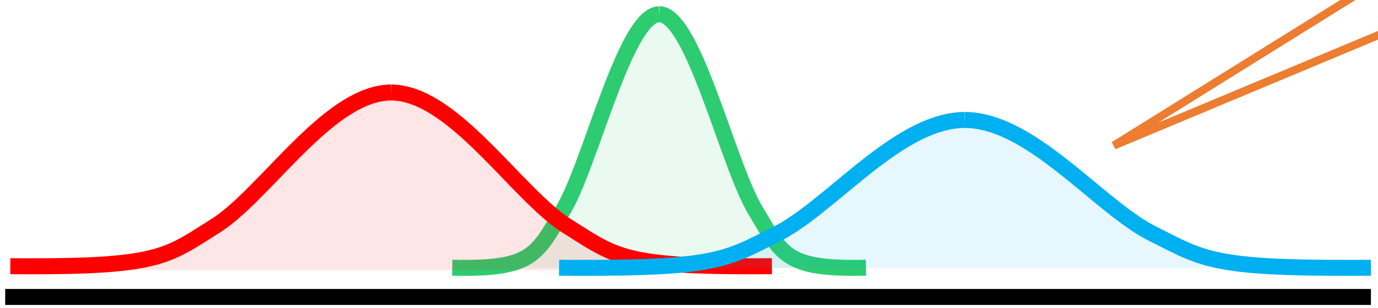


The Naïve Bayes model



“Mixture” of
3 Gaussians

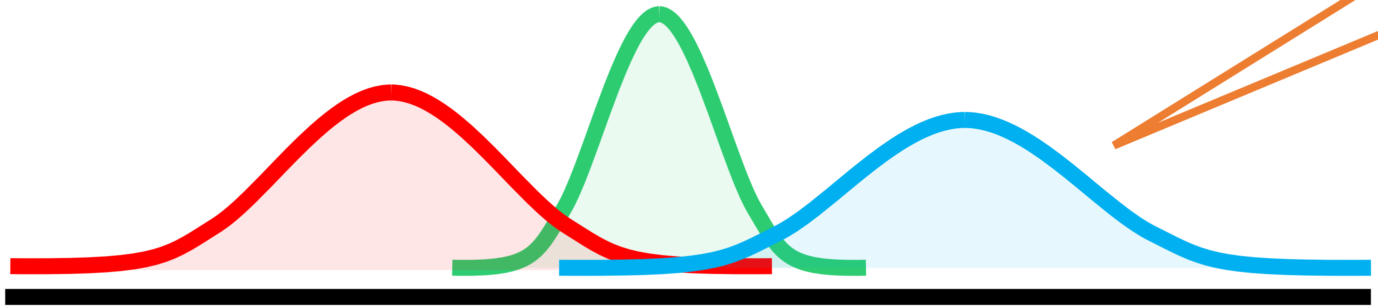
The Naïve Bayes model



"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}$,
$\mathrm{y} \in \{\color{red}{\mathbf{1}}, \color{green}{\mathbf{2}}, \color{cyan}{\mathbf{3}}\}$

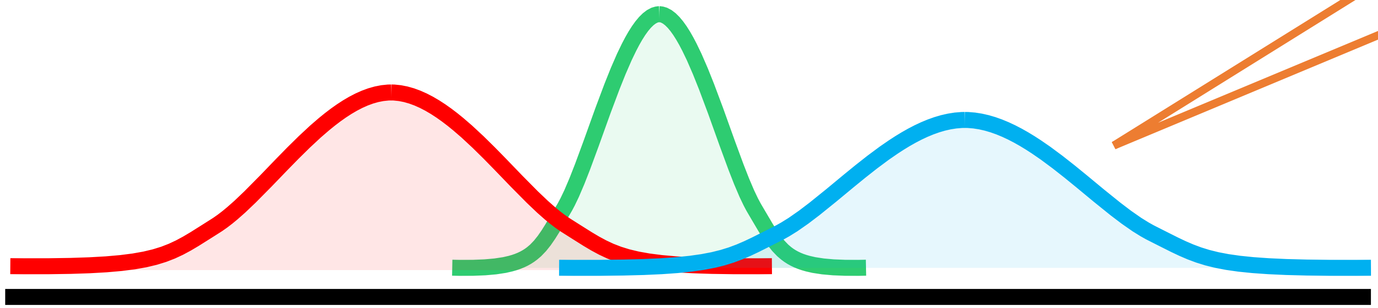
The Naïve Bayes model



"Mixture" of
3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$\mathrm{y} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$

The Naïve Bayes model

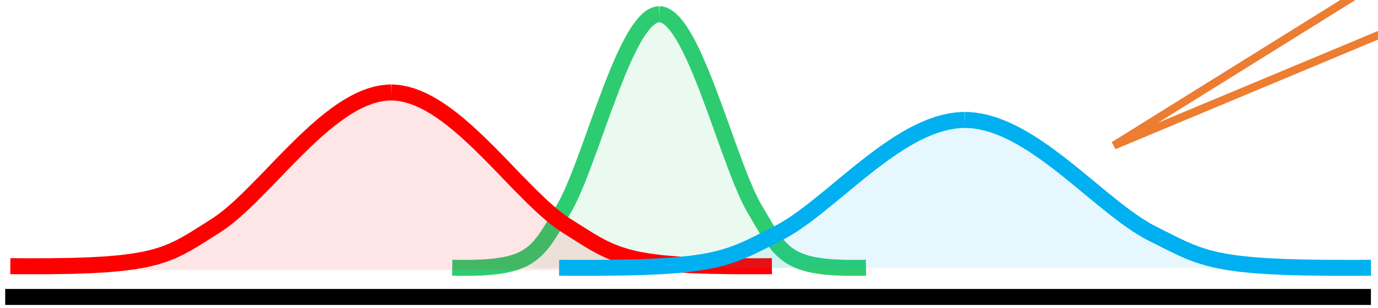


"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

$$\mathbb{P}[\mathbf{x}, y] = \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x} \mid y]$$

The Naïve Bayes model



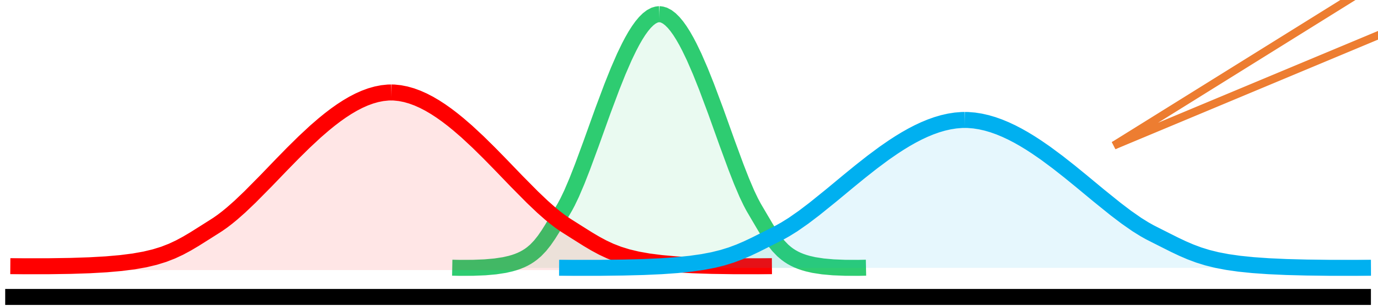
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$\mathbb{P}[\mathbf{x}, y] = \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x} \mid y]$$

The Naïve Bayes model



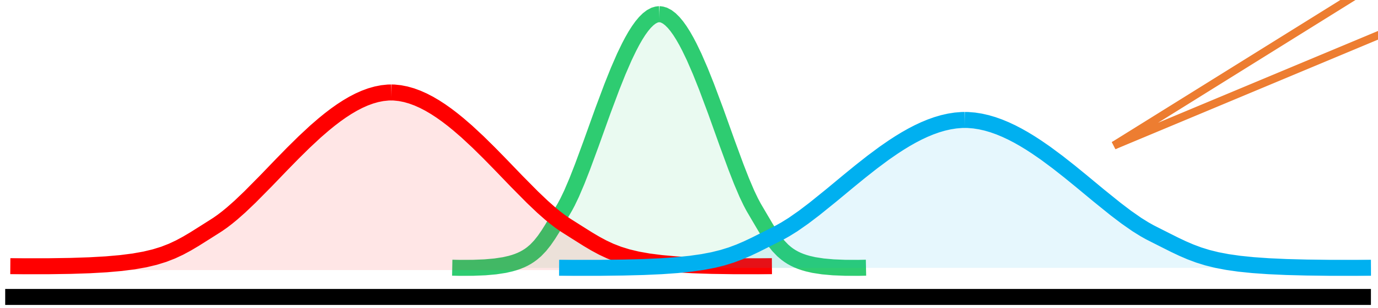
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$\mathbb{P}\left[\mathbf{x}, y\right] = \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x} \,|\, y\right]$$
$$= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y\right]$$

The Naïve Bayes model



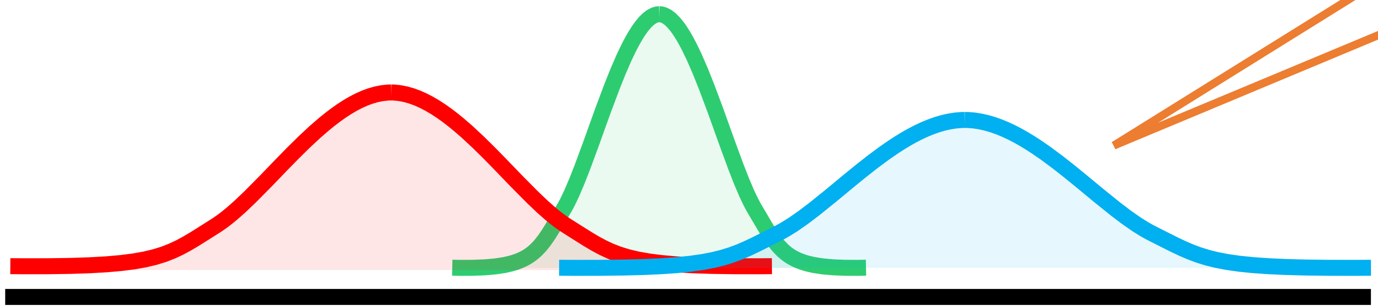
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$
\begin{aligned}
\mathbb{P}\left[\mathbf{x}, y\right] &= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x} \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right]
\end{aligned}
$$

The Naïve Bayes model



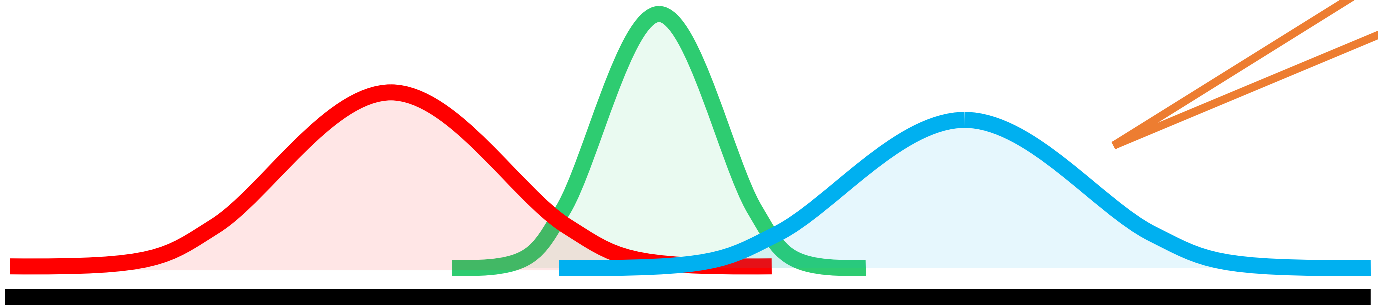
"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$
\begin{aligned}
\mathbb{P}\left[\mathbf{x}, y\right] &= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x} \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right]
\end{aligned}
$$

The Naïve Bayes model



"Mixture" of 3 Gaussians

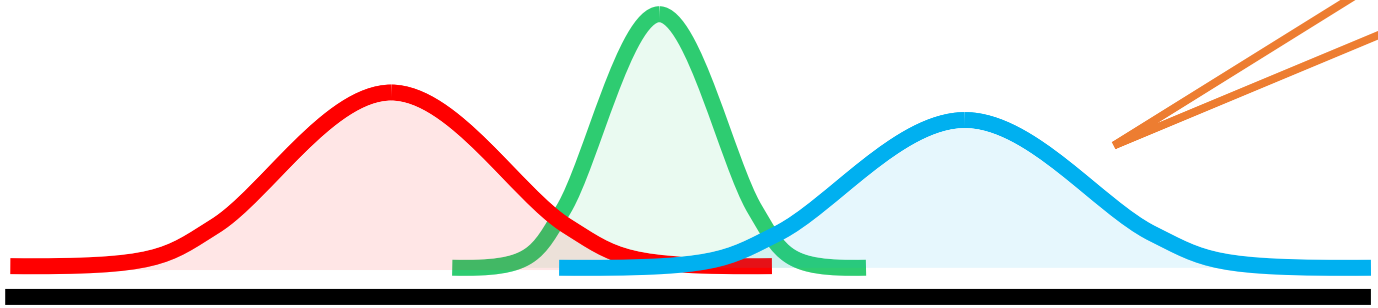
$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$\begin{aligned}
\mathbb{P}[\mathbf{x}, y] &= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x} \,|\, y] \\
&= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y] \\
&= \mathbb{P}[y] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x}_1 \,|\, y] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right]
\end{aligned}$$

Naïve Bayes assumption
$\mathbb{P}[x_j | \, x_k, y] = \mathbb{P}[x_j | y]$ if $j \neq k$

The Naïve Bayes model



"Mixture" of 3 Gaussians

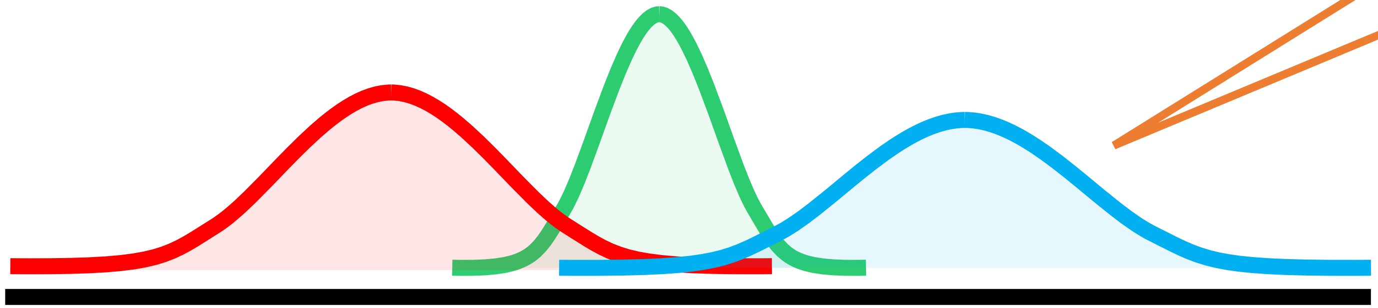
$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

$$
\begin{aligned}
\mathbb{P}[\mathbf{x}, y] &= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x} \mid y] \\
&= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \mid y] \\
&= \mathbb{P}[y] \cdot \mathbb{P}\left[\mathbf{x}_1 \mid \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \mid y\right] \\
&= \mathbb{P}[y] \cdot \mathbb{P}[\mathbf{x}_1 \mid y] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \mid y\right] \\
&= \ldots
\end{aligned}
$$

Chain rule of probabilty

Naïve Bayes assumption
$\mathbb{P}[x_j | x_k, y] = \mathbb{P}[x_j | y]$ if $j \neq k$

The Naïve Bayes model



"Mixture" of 3 Gaussians

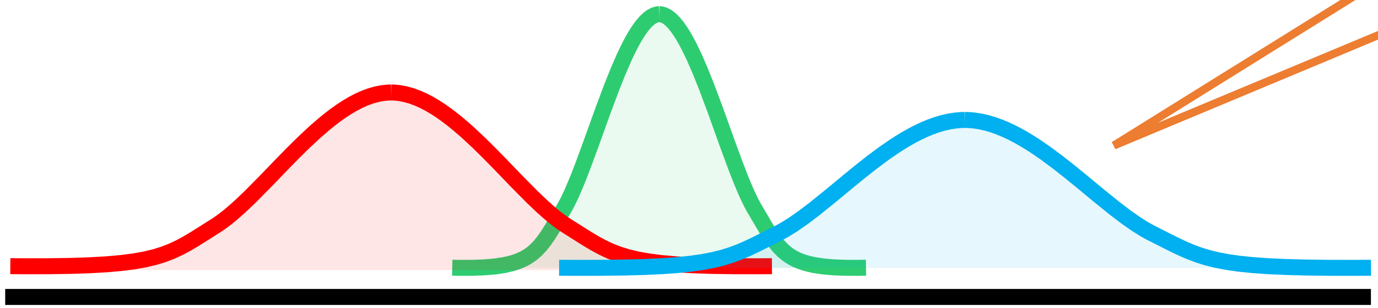
$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$
\begin{aligned}
\mathbb{P}\left[\mathbf{x}, y\right] &= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x} \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \ldots \\
&= \mathbb{P}\left[y\right] \cdot \prod_{j=1}^{d} \mathbb{P}\left[\mathbf{x}_i \,|\, y\right]
\end{aligned}
$$

Naïve Bayes assumption
$\mathbb{P}[x_j | \, x_k, y] = \mathbb{P}[x_j | y]$ if $j \neq k$

The Naïve Bayes model



"Mixture" of 3 Gaussians

$\boldsymbol{x} \in \mathbb{R}^d$,
$y \in \{1, 2, 3\}$

Chain rule of probabilty

$$
\begin{aligned}
\mathbb{P}\left[\mathbf{x}, y\right] &= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x} \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \mathbb{P}\left[y\right] \cdot \mathbb{P}\left[\mathbf{x}_1 \,|\, y\right] \cdot \mathbb{P}\left[\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}^d \,|\, y\right] \\
&= \ldots \\
&= \mathbb{P}\left[y\right] \cdot \prod_{j=1}^{d} \mathbb{P}\left[\mathbf{x}_i \,|\, y\right]
\end{aligned}
$$

Naïve Bayes assumption
$\mathbb{P}[x_j |\, x_k, y] = \mathbb{P}[x_j | y]$ if $j \neq k$

Already seen how to model in 1D

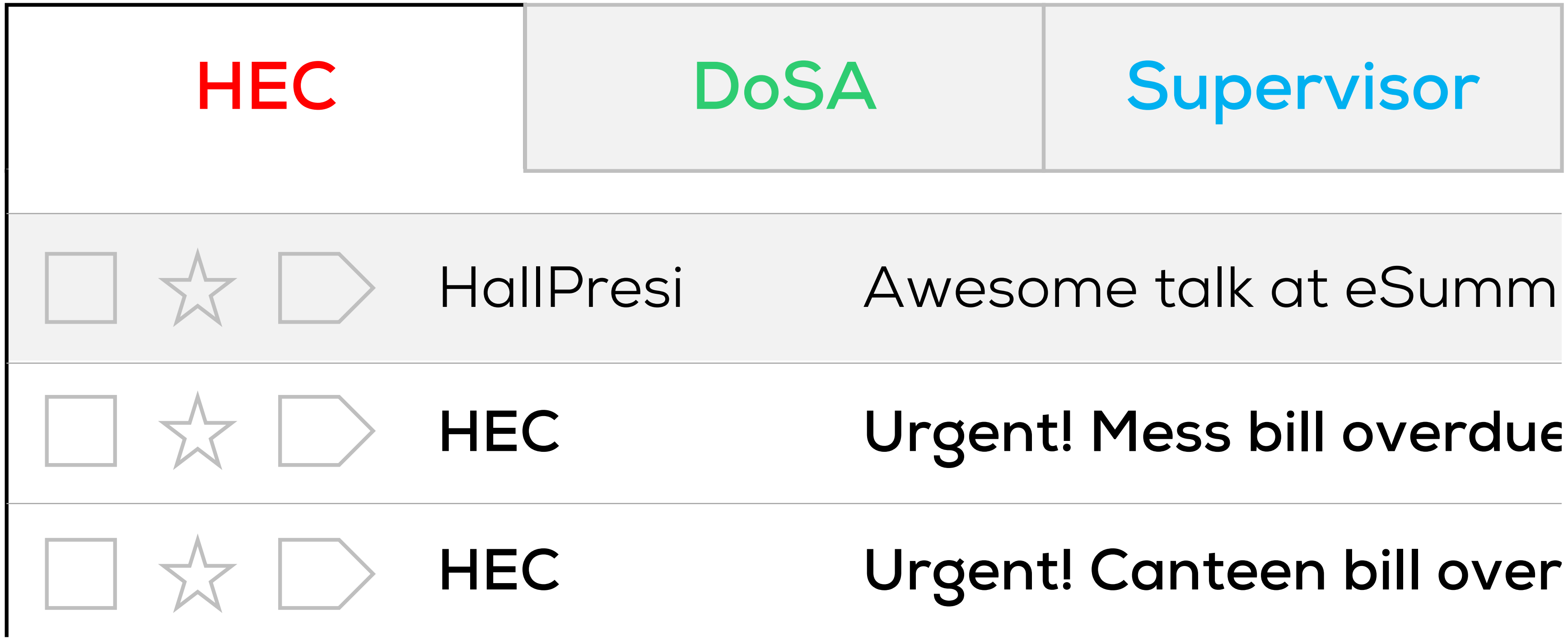
App: Email Categorizer using Naïve Bayes

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

App: Email Categorizer using Naïve Bayes



$\mathbf{x}$

talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

$\mathbf{x}$

talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

$$y \in \{ {\color{red}\bullet}, {\color{green}\bullet}, {\color{cyan}\bullet} \}$$

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

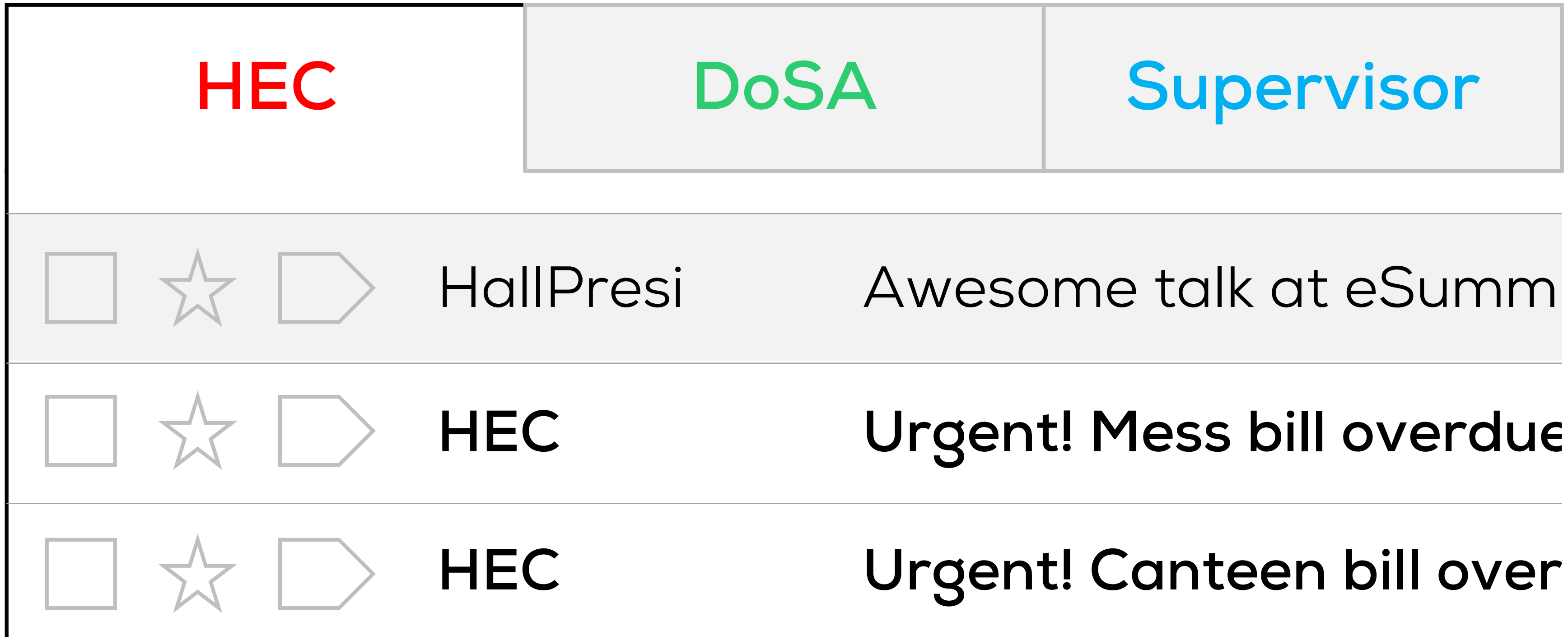
Bag of words feature can record just occurrence or count

$\mathbf{x}$

talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

$y \in \{ \bullet , \bullet , \bullet \}$

App: Email Categorizer using Naïve Bayes



Bag of words feature can record just occurrence or count

$\mathbf{x}$

1	0	0	0	0	0	0	1	0	0	0	0	0
talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

$y \in \{$ 🔴 , 🟢 , 🔵 $\}$

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

Sender	Subject
HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

Choice of words crucial – stemming, throw away articles etc

Bag of words feature can record just occurrence or count

$\mathbf{x}$

1	0	0	0	0	0	0	1	0	0	0	0	0
talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

$$y \in \{ \bullet , \bullet , \bullet \}$$

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

Commonly used in NLP

Choice of words crucial – stemming, throw away articles etc

Bag of words feature can record just occurrence or count

$\mathbf{x}$

1	0	0	0	0	0	0	1	0	0	0	0	0
talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report

$$y \in \{ {\color{red}\bullet}, {\color{green}\bullet}, {\color{cyan}\bullet} \}$$

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

Commonly used in NLP

Choice of words crucial – stemming, throw away articles etc

Bag of words feature can record just occurrence or count

$\mathbf{x}$

talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report
1	0	0	0	0	0	0	1	0	0	0	0	0

Usually very high dimensional

$y \in \{ \bullet , \bullet , \bullet \}$

App: Email Categorizer using Naïve Bayes

HEC	DoSA	Supervisor

HallPresi	Awesome talk at eSumm
HEC	**Urgent! Mess bill overdue**
HEC	**Urgent! Canteen bill over**

Commonly used in NLP

Choice of words crucial – stemming, throw away articles etc

Bag of words feature can record just occurrence or count

$\mathbf{x}$

talk	meet	project	wake	meeting	wallet	keys	awesome	lost	trip	lost	mess	report
1	0	0	0	0	0	0	1	0	0	0	0	0

$y \in \{ \quad , \quad , \quad \}$

Usually very high dimensional

Usually very very sparse

App: Email Categorizer using Naïve Bayes

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\bullet] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\bullet] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid \bullet] = \frac{|\#\text{emails from sup. with "awesome"}|}{|\#\text{total emails from supervisor}|}$$

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\bullet] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid \bullet] = \frac{|\#\text{emails from sup. with "awesome"}|}{|\#\text{total emails from supervisor}|}$$

At test time …

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\bullet] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid \bullet] = \frac{|\#\text{emails from sup. with "awesome"}|}{|\#\text{total emails from supervisor}|}$$

At test time …

$$\mathbb{P}[\mathbf{x}^t, \bullet] = \mathbb{P}[\bullet] \cdot \prod_{j=1}^{d} \mathbb{P}[\mathbf{x}_i^t \mid \bullet]$$

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\bullet] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid \bullet] = \frac{|\#\text{emails from sup. with "awesome"}|}{|\#\text{total emails from supervisor}|}$$

$$\mathbb{P}[awesome = 0|\bullet]$$
$$= 1 - \mathbb{P}[awesome = 1|\bullet]$$

At test time …

$$\mathbb{P}[\mathbf{x}^t, \bullet] = \mathbb{P}[\bullet] \cdot \prod_{j=1}^{d} \mathbb{P}[\mathbf{x}_i^t \mid \bullet]$$

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[{\color{cyan}\bullet}] = \frac{|\#\text{emails from supervisor}|}{|\#\text{total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid {\color{cyan}\bullet}] = \frac{|\#\text{emails from sup. with "awesome"}|}{|\#\text{total emails from supervisor}|}$$

$$\mathbb{P}[awesome = 0 | {\color{cyan}\bullet}]$$
$$= 1 - \mathbb{P}[awesome = 1 | {\color{cyan}\bullet}]$$

At test time ...

$$\mathbb{P}[\mathbf{x}^t, {\color{cyan}\bullet}] = \mathbb{P}[{\color{cyan}\bullet}] \cdot \prod_{j=1}^{d} \mathbb{P}[\mathbf{x}_i^t | {\color{cyan}\bullet}]$$

$$\hat{y}^t = \arg\max\{\mathbb{P}[\mathbf{x}^t, {\color{red}\bullet}], \mathbb{P}[\mathbf{x}^t, {\color{green}\bullet}], \mathbb{P}[\mathbf{x}^t, {\color{cyan}\bullet}]\}$$

App: Email Categorizer using Naïve Bayes

$$\mathbb{P}[\color{#00B0F0}{\bullet}] = \frac{|\text{\#emails from supervisor}|}{|\text{\#total emails}|}$$

$$\mathbb{P}[\text{awesome} = 1 \mid \color{#00B0F0}{\bullet}] = \frac{|\text{\#emails from sup. with "awesome"}|}{|\text{\#total emails from supervisor}|}$$

$\mathbb{P}[awesome = 0|\color{#00B0F0}{\bullet}]$
$= 1 - \mathbb{P}[awesome = 1|\color{#00B0F0}{\bullet}]$

At test time …

$$\mathbb{P}[\mathbf{x}^t, \color{#00B0F0}{\bullet}] = \mathbb{P}[\color{#00B0F0}{\bullet}] \cdot \prod_{j=1}^{d} \mathbb{P}\left[\mathbf{x}_i^t | \color{#00B0F0}{\bullet}\right]$$

Will give the same result as
$\arg\max\{\mathbb{P}[\color{red}{\bullet} | \mathbf{x}^t], \mathbb{P}[\color{green}{\bullet} | \mathbf{x}^t], \mathbb{P}[\color{#00B0F0}{\bullet} | \mathbf{x}^t]\}$

$$\hat{y}^t = \arg\max\{\mathbb{P}[\mathbf{x}^t, \color{red}{\bullet}], \mathbb{P}[\mathbf{x}^t, \color{green}{\bullet}], \mathbb{P}[\mathbf{x}^t, \color{#00B0F0}{\bullet}]\}$$

App: Automatic Email Generator!

Class proportions

- Choose a category from {HEC, DoSA, Supervisor}
  - Toss a 3-sided "coin" aka categorical/multinoulii distribution using $\mathbb{P}[\bullet]$
  - Say we chose HEC
- For each word in your dictionary of d words, toss a Bernoulli coin to decide whether to include that word in the mail or not
  - For $j \in [d]$, toss a coin that lands heads with probability $\mathbb{P}\left[x_j = 1 \mid \bullet\right]$
- Collect all words for which the toss landed heads
- Compose an email using only those words (and maybe a few articles, prepositions etc)
- Congratulations, you can now ask your HEC to stop sending you emails – you will generate them yourself!

Already learnt from training data!

Please give your Feedback

http://tinyurl.com/ml17-18afb