Parameter Estimation - II

EE698V - Machine Learning for Signal Processing

Vipul Arora

Recap

- We saw ML and MAP estimates
- For ML estimate:
  - ML estimate of f(~~x~~ θ) = f(ML estimate of ~~x~~ θ)
  - ML estimate of σ^2 = (ML estimate of σ)^2

- samples
- decide a model
- $\mathcal{L} = p(\text{samples}; \theta)$

$$\theta_{ML} = \arg\max_{\theta} p(s; \theta)$$

$$\theta_{MAP} = \arg\max_{\theta} p(\theta | s)$$

$$= \arg\max_{\theta} p(s|\theta) \underbrace{p(\theta)}_{\text{prior}}$$

$$\theta : \quad \frac{\partial \ln \mathcal{L}}{\partial \theta} = 0 \equiv \frac{\partial \ln \mathcal{L}}{\partial f} \frac{\partial f}{\partial \theta} = 0$$

Sufficient Statistics and Sequential Estimation

Reference: PRML Section 2.3.4-5

Sufficient Statistics

- $\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N} s_i$
- Calculation of μ_{ML} needs only sum of samples and N
- If data arriving sequentially, and we want to update the estimate, we need not store all the samples, but just the sufficient statistics
- How would you update μ_{ML} with s_t arriving at every instant t?

Sequential Estimation of μ_{ML}

- I have $\mu_{\mathrm{ML}}^{(N-1)}$ and a new sample s_N arrives

- $\mu_{\mathrm{ML}}^{(N)} = \mu_{\mathrm{ML}}^{(N-1)} + \frac{1}{N}\left(s_N - \mu_{\mathrm{ML}}^{(N-1)}\right)$ $\quad \|$

$$\mu_{ML}^{(N)} = \frac{1}{N}\sum_{t=1}^{N} s_t \qquad \mu_{ML}^{(N-1)} = \frac{1}{N-1}\sum_{t=1}^{N-1} s_t$$

$$= \frac{1}{N}\left(\underbrace{\sum_{t=1}^{N-1} s_t}_{(N-1)\,\mu_{ML}^{(N-1)}} + s_N\right)$$

Sequential Estimation of σ^2_{ML}

- $\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N}(s_i - \mu)^2$

ML

- DIY ...

$\underbrace{\sum_{t=1}^{N} s_t^2}$, $\underbrace{\sum_{t=1}^{N} s_t}$

$E[(x-\mu)^2] = E[x^2 - 2\mu x + \mu^2]$

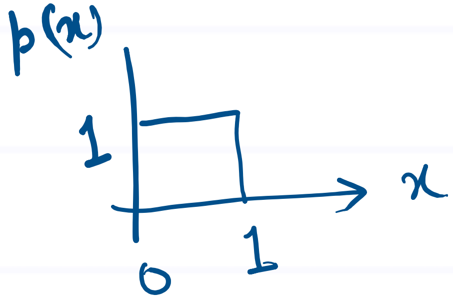
$= E[x^2] - 2\underbrace{E[\cancel{\mu} x]}_{\mu}\mu + \mu^2$

$= E[x^2] - \mu^2$

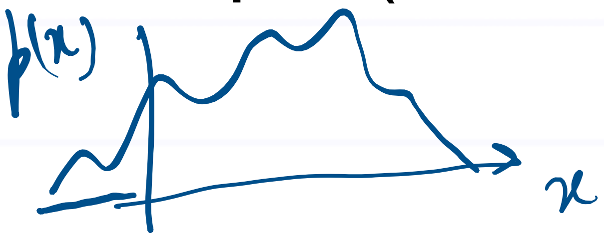
Sampling

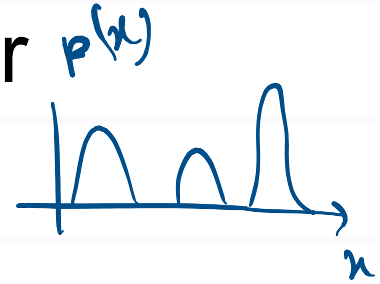
How to sample from a given pdf

- Programming platforms have random number generators



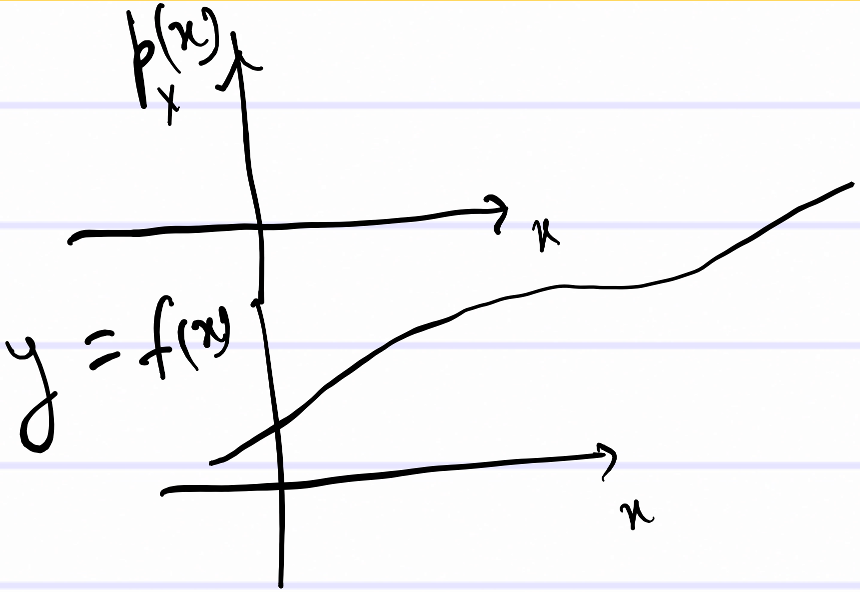
- Generally, uniform(0,1) is available
- How to obtain samples from any general pdf (scalar r.v.)?

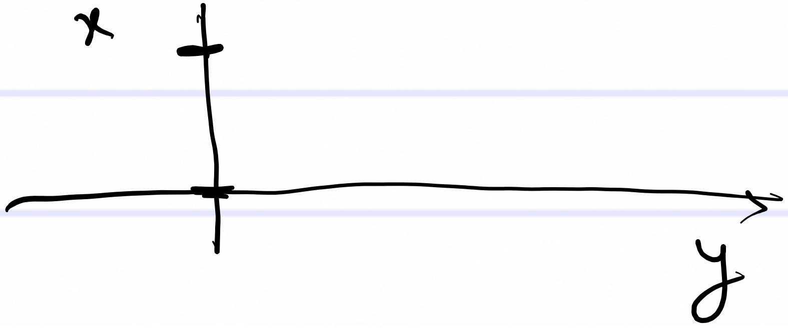




- Hint: use transformation of variables

How to sample from a given pdf





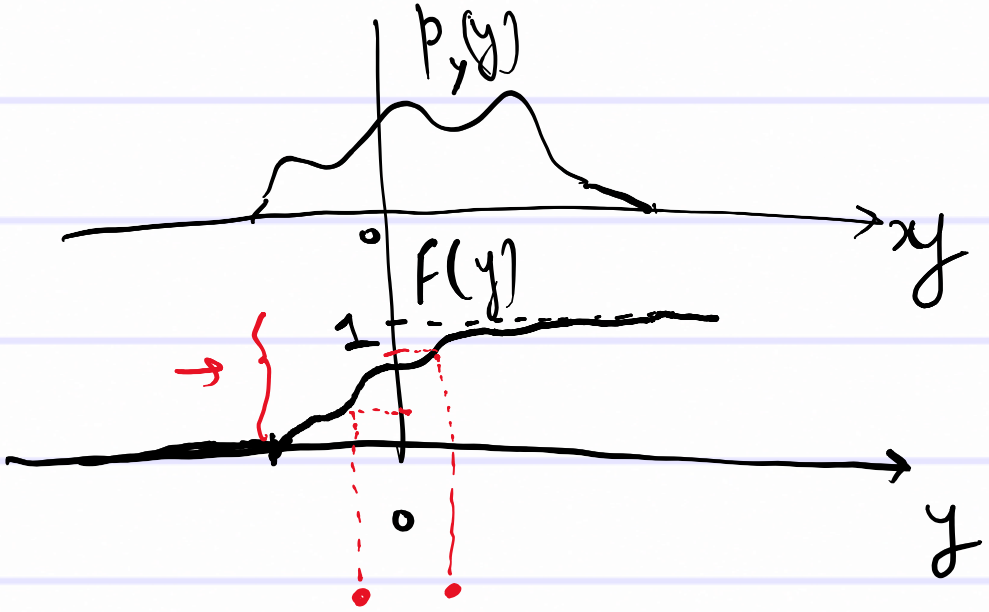
$$p_Y(y) = \frac{p_X(x)}{\left|\frac{dy}{dx}\right|}$$

$$p_Y(y) = \frac{p_X(x)}{\left|dy/dx\right|} \longrightarrow \text{uniform } (0,1)$$

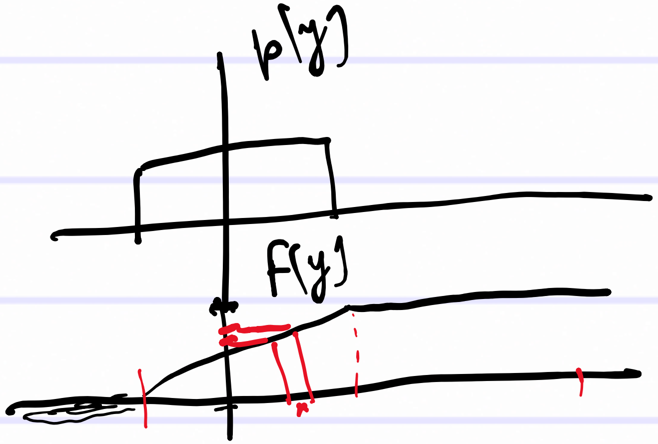
given non std.

$$= 1 \left|\frac{dx}{dy}\right| \qquad x \in (0,1)$$

Take $y = CDF^{-1}(x)$ $\qquad x = CDF(y)$



$$F(y) = \int_{-\infty}^{y} p_y(y')\, dy'$$



Gaussian Distribution

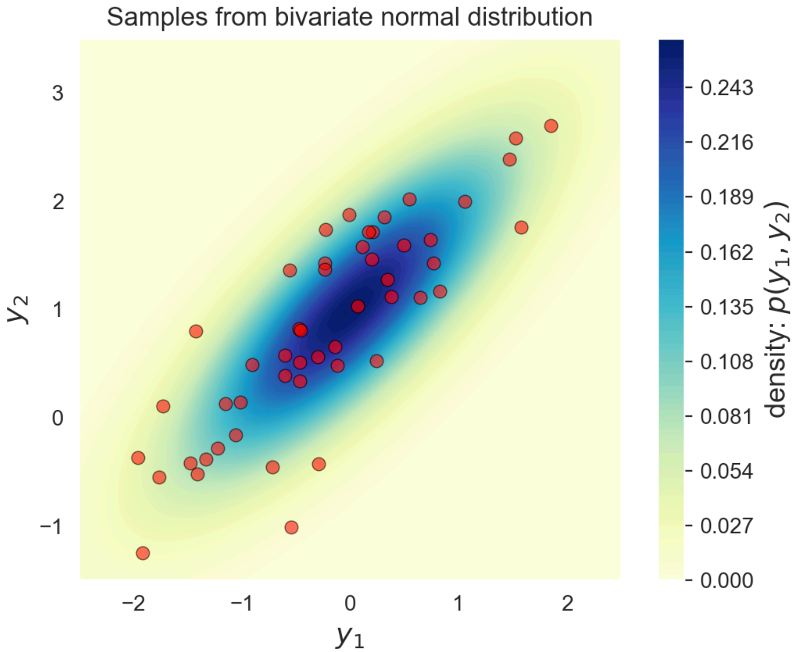
Reference: PRML Section 2.3 intro

Multivariate Gaussian pdf

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \cdots \\ (x_2, x_1) & \\ \vdots & \end{bmatrix}$$

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ \underbrace{-\frac{1}{2}(\overset{D\times 1}{\mathbf{x}} - \overset{D\times 1}{\boldsymbol{\mu}})^{\mathrm{T}} \overset{D\times D}{\boldsymbol{\Sigma}}^{-1} (\overset{D\times 1}{\mathbf{x}} - \boldsymbol{\mu})}_{\text{scalar } 1\times 1} \right\}$$

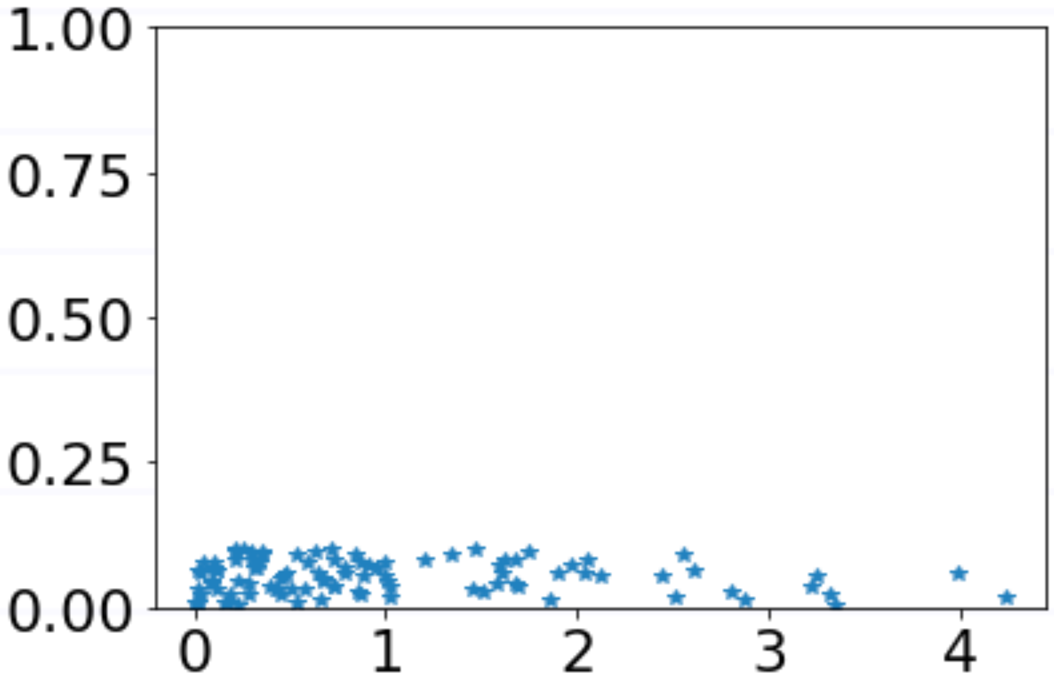


- $\boldsymbol{\mu}_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{s}_i$
- $\Sigma_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{s}_i - \boldsymbol{\mu}_{\mathrm{ML}})(\boldsymbol{s}_i - \boldsymbol{\mu}_{\mathrm{ML}})^T$

$\mathrm{cov}(x_i, x_j) = 0$ if $i \neq j$

if x_i's are indep.

How would you model this data?

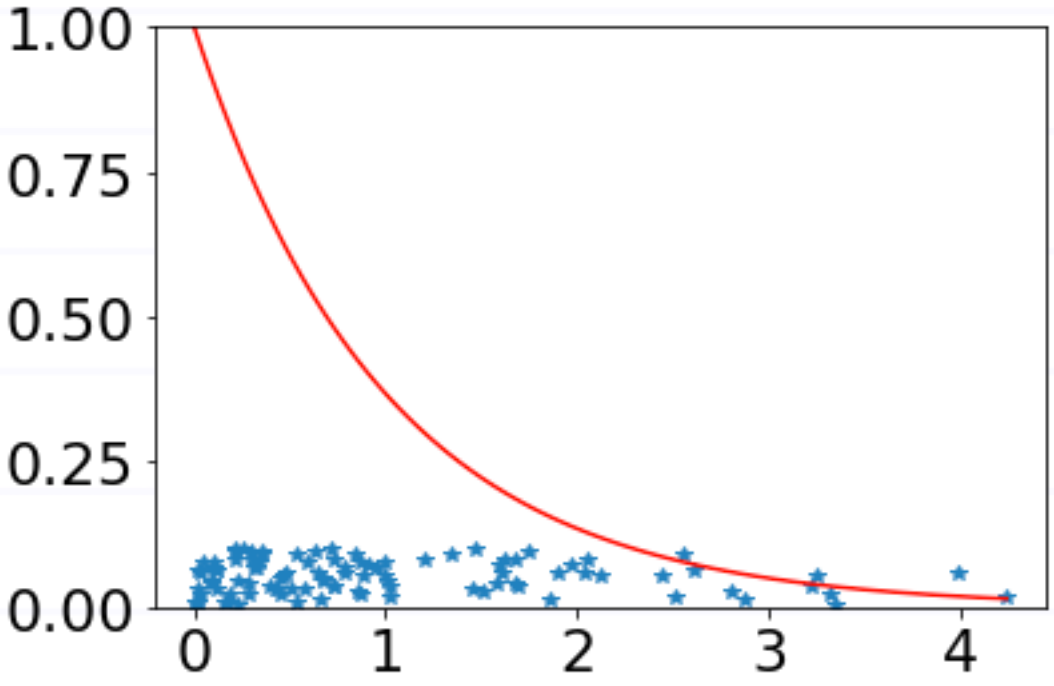


if I estimate these samples with a unif.(0, 4.2)

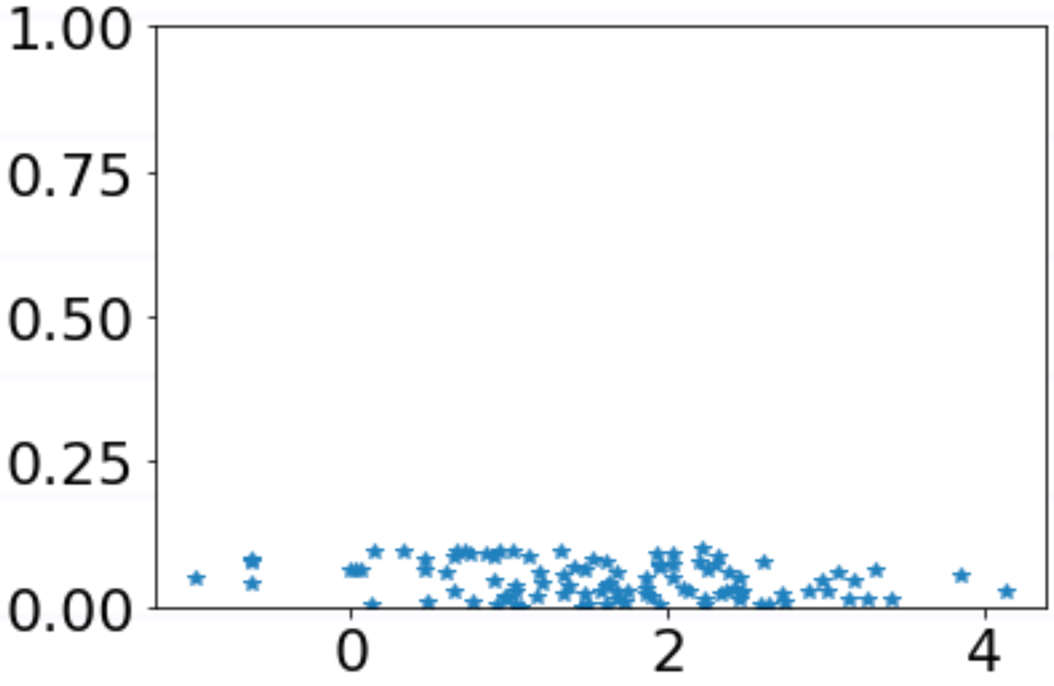
K.L. ( ) ↑

Exponential pdf

$$p(x;\ \beta) = \frac{1}{\beta} e^{-x/\beta};\, x \geq 0, \beta > 0$$

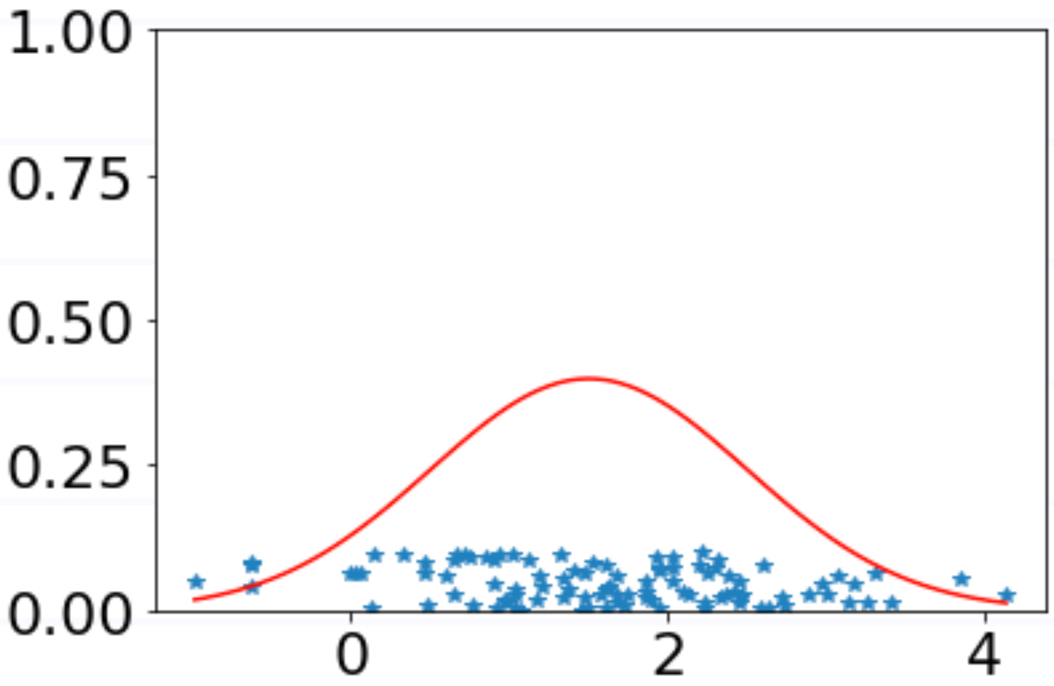


How would you model this data?

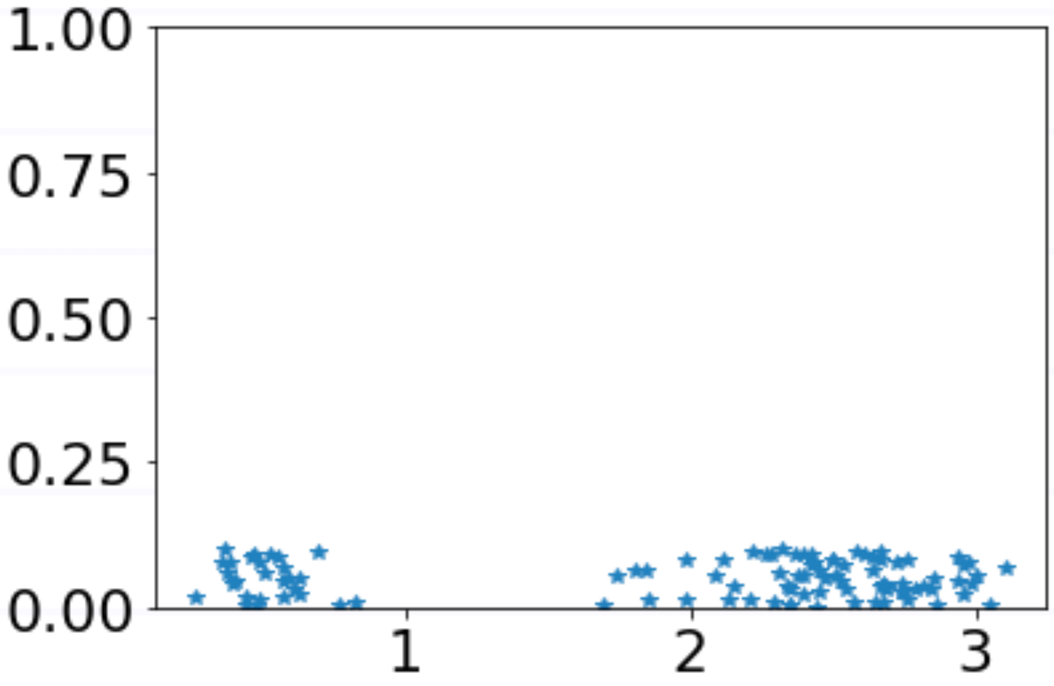


Normal or Gaussian pdf

$$p(x;\ \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2};\ \sigma > 0$$



How would you model this data?



Latent Variable Models

Reference: PRML Chapter 9

Latent Variables

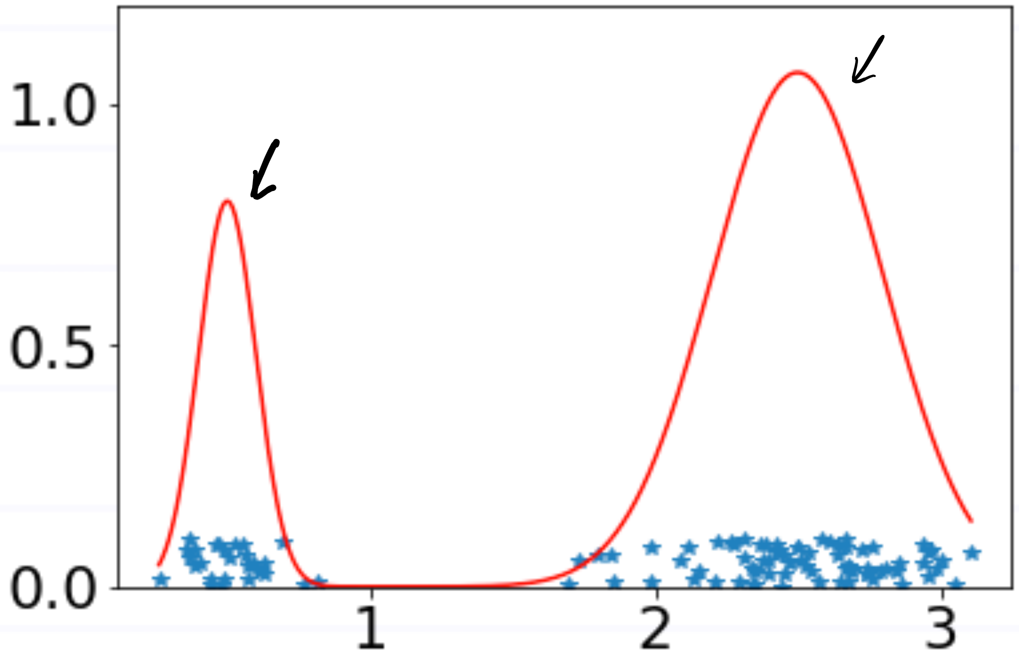
- Latent variables help complicated distributions to be formed from simpler components
- $p(x) = p(x|z=1)p(z=1) + p(x|z=2)p(z=2)$
- $p(x) = \sum_z \underbrace{p(x|z)}_{\text{gaus}} \underbrace{p(z)}_{\text{p.m.f.}}$

$\sum_z p(x, z)$

discrete r.v. $\in \{1, 2\}$

Gaussian Mixture Model

$$p(x) = \sum_k p(z_k)\mathcal{N}(x;\mu_k,\sigma_k)\,;\, z_k \in \{0,1\}, \sum_k z_k = 1$$



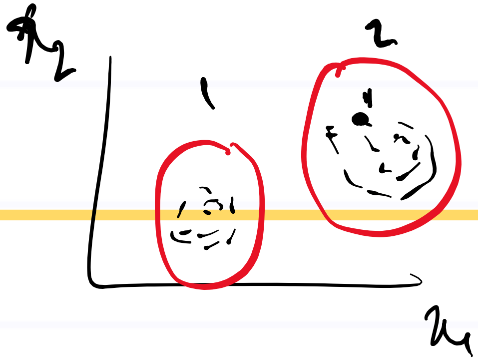
$Z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

one hot vector

K-means Clustering

- To understand GMMs, let us begin with this
- Identifying clusters in data samples
- **Reference**: PRML Section 9.1

K-means Clustering

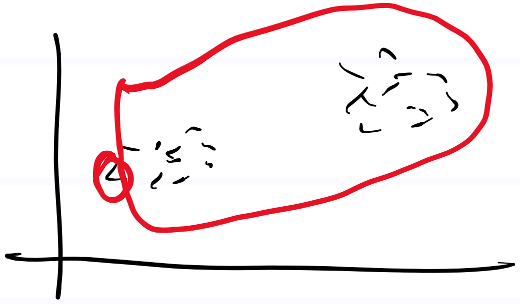


- Distortion measure: $\mathcal{L} = \sum_{n,k} r_{nk} \|x_k - \mu_k\|^2$
- k is the cluster index
- n is the sample index
- $r_{nk} = 1$ if sample n belongs to cluster k (hard clustering)
- $\mathcal{L}$ avoids very large (+ singular) clusters

$r_{11} = 0$

$r_{12} = 1$

membership fn



Optimal value of r_{nk} and μ_k

$$\mathcal{L} = \sum_{n,k} r_{nk} \|x_k - \mu_k\|^2$$

Given r_{nk} $\quad \frac{\partial \mathcal{L}}{\partial \mu_{k'}} = 2 \sum_n r_{nk'} (x_{k'} - \mu_{k'}) = 0 \Rightarrow \mu_{k'} = \frac{\sum_n r_{nk'} x_{k'}}{\sum_n r_{nk'}}$

Given $\mu_{k'}$, $\quad r_{nk} = \begin{cases} 1, & \text{if } k = \underset{k'}{\operatorname{argmin}} \|x_n - \mu_{k'}\| \\ 0, & \text{o.w.} \end{cases}$

K-means iteration