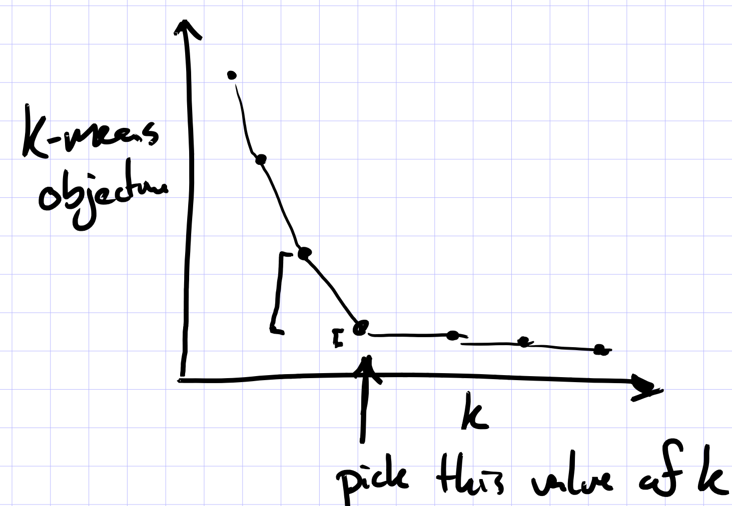
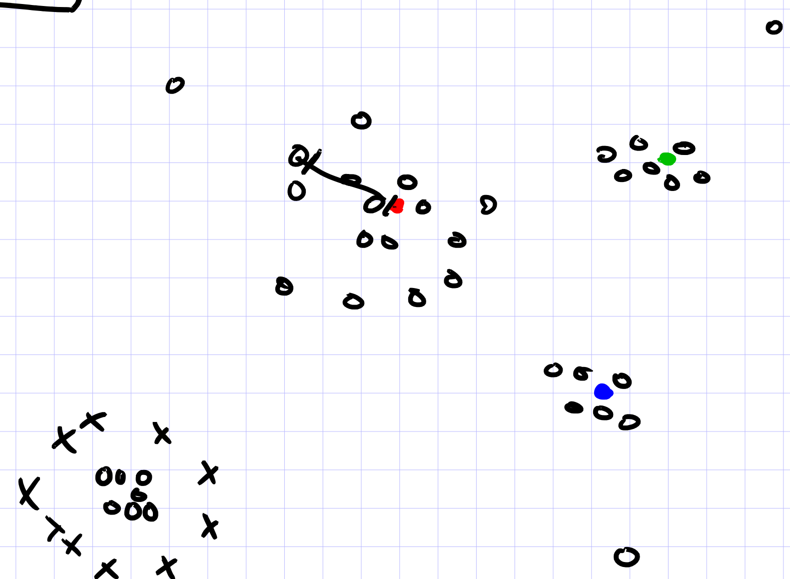
Lecture 16 : 3/20/17

How to choose k?





Background for EM/GMM

Gaussian Distribution

Case #1 : Univariate (1D)

parameters : $\mu \in \mathbb{R}$ $\quad \sigma^2 \in \mathbb{R}^+$

support : $x \in \mathbb{R}$

pdf: $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$

pdf if $\sigma^2 = 1$: $f(x \mid \mu, \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\}$

Case #2 : Multivariate

parameters : $\vec{\mu} \in \mathbb{R}^M$ $\quad \Sigma \in \mathbb{R}^{M \times M}$ positive-definite matrix

support : $\vec{x} \in \mathbb{R}^M$

pdf : $f(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^M |\Sigma|}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\mu)\right\}$

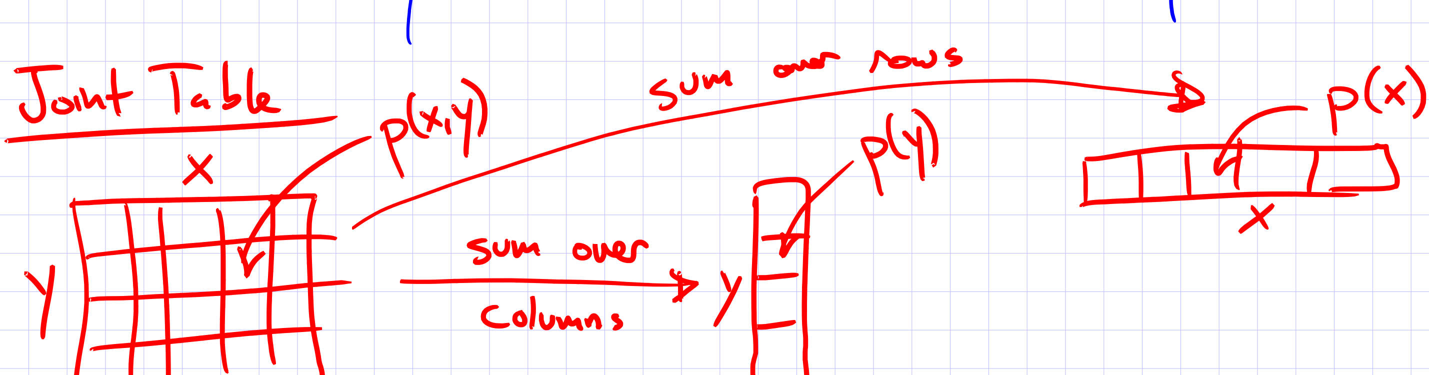
determinant

pdf if $\Sigma = I$: $f(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^M}} \exp\left\{-\frac{1}{2}\|\vec{x}-\mu\|^2\right\}$

Joint vs. Marginal

Let X and Y be r.v.s where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

	Discrete (pmf)	Continuous (pdf)
joint dist. of x and y	$p(x,y) = Pr[X=x, Y=y]$	$f(x,y)$
marginal dist. of x	$p(x) = \sum_{y \in \mathcal{Y}} p(x,y)$	$f(x) = \int_y f(x,y)\,dy$



Gaussian Mixture Model (GMM)

What do we need for probabilistic approach to learning?

1. Data
2. Model (Generative Story)
3. Decision Rule
4. Objective (eg. likelihood, conditional likelihood)
5. Optimization Method (e.g. closed-form, Gradient Descent)

Gaussian Naive Bayes	Gaussian Discriminant Analysis	Gaussian Mixture Model
$\mathcal{D} = \{(\vec{x}^{i}, y^{i})\}_i$ Supervised	$\mathcal{D} = \{(\vec{x}^{i}, y^{i})\}_i$ Supervised	$\mathcal{D} = \{\vec{x}^{i}\}_i$ Unsupervised
$y \sim \text{Categorical}(\phi)$ $x_1 \sim \text{Gaussian}(\mu_{1y}, \sigma^2_{1y})$ $\vdots$ $x_M \sim \text{Gaussian}(\mu_{My}, \sigma^2_{My})$	$y \sim \text{Categorical}(\phi)$ $\vec{x} \sim \underbrace{\text{Gaussian}(\vec{\mu}_y, \Sigma_y)}$	$z \sim \text{Categorical}(\phi)$ $\vec{x} \sim \underbrace{\text{Gaussian}(\vec{\mu}_z, \Sigma_z)}$
$p(\vec{x}, y) = p(y) \prod_{m=1}^{M} p(x_m \mid y)$	$p(\vec{x}, y) = p(y)\, p(\vec{x} \mid y)$	$p(\vec{x}) = \sum_z p(\vec{x}, z) = \sum_z p(z) p(x \mid z)$
$\ell(\phi, \mu, \sigma^2) = \sum_{i=1}^{N} \log p(\vec{x}^{(i)}, y^{(i)})$ (log-likelihood)	$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{N} \log p(\vec{x}^{(i)}, y^{(i)})$ (log-likelihood)	$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{N} \log p(\vec{x}^{(i)})$ $= \sum_{i=1}^{N} \log \sum_z p(\vec{x}^{(i)}, z)$ (log marginal likelihood)

multivariate

generative story is identical!!

$\boxed{\text{How would you train GDA}}$

Optimize $\ell(\phi, \mu, \Sigma)$ in closed-form (set derivatives equal to 0 and solve)

Counting-style algorithm.

$\boxed{\text{How would you train GMM?}}$

Not so easy... the summation over z couples the parameter