Lecture 10

*) M.V.G. MLE / Bernoulli & M.V.

*) MAP

MAP estimate:

$$\max_{\theta} \left[\sum_{i=1}^{N} \ln p(x \mid \theta)\right] + \ln p(\theta)$$

$$\max_{\theta} \left[\sum_{i=1}^{N} \left\{x_i \ln \theta + (1-x_i) \ln (1-\theta)\right\}\right] + (\alpha-1) \ln \theta + (\beta-1) \ln (1-\theta)$$

$$p(\theta) \prod_{i=1}^{N} p(x \mid \theta)$$

$$p(\theta)\, \underbrace{p(1 \mid \theta)}_{\prod_{i=1}^{N}}$$

$$\theta_{MAP} = \frac{\sum_i x_i + \alpha - 1}{N + \alpha + \beta - 2}$$

$$\theta_{MLE} = \frac{\sum_i x_i}{N}$$

N iid's. $x_1, x_2 \cdots x_4$ $N=4$

$$ABC = C(AB)$$

$$\overset{H}{x_1 = 1}, \quad \overset{T}{x_2 = 0}, \quad \overset{T}{x_3 = 0}, \quad \overset{T}{x_4 = 0}$$

$$\theta_{MLE} = \frac{1}{4}$$

$\alpha = 5, \beta = 5$

$$\theta_{MP} = \frac{1+\alpha-1}{4+\alpha+\beta-2} = \frac{5}{12}$$

Demo

- Once you have ML/MAP estimates, you could plug them in likelihood function for a classification setting.
- Take an eg of two category case with 3-d and 4 samples.

ω_1

$$p(\omega|x) = p(x|\theta)\,p(\omega)$$

MLE/MAP

	SML	GA	Post
	d_1	d_2	d_3
x_1	1	0	1
x_2	0	0	0
x_3	0	0	1
x_4	1	0	1

$$p(x|\theta) = \theta_1^{x_1}(1-\theta_1)^{1-x_1} \cdot \theta_2(1-\theta_2)^{1-x_2} \cdot \theta_3^{x_3}(1-\theta_3)^{1-x_3}$$

$\theta_{1ML} = \frac{1}{2}$, $\theta_{2ML} = 0$, $\theta_{3ML} = 3/4$

ω_2, $\quad d_1 \quad d_2 \quad d_3 \qquad \theta_1 = \frac{1}{4}, \; \theta_2 = \frac{2}{4}, \; \theta_3 = \frac{1}{4}$

$x_1 \quad 1 \quad 0 \quad 0$

x_2

x_3

x_4

$$P(x|\theta) = (\theta_1)^{x_1} (1-\theta_1)^{1-x_1} \cdot \theta_2 \cdot \theta_3$$

$$P(\omega_1) = P(\omega_2) = 1/2$$

$X_0 \rightarrow$ test sample $\rightarrow \{1, 1, 1\}$ $\nearrow \omega_1$? $\searrow \omega_2$?

$$P(\omega_1|x) \propto P(x|\cancel{\theta}) P(\omega_1) = P(x|\omega_1) P(\omega_1)$$

$$P(\omega_2|x) \propto P(x|\cancel{\theta}) P(\omega_2) = P(x|\omega_2) P(\omega_2)$$

$$P(\omega_1|x) \equiv P(x|\omega_1) P(\omega_1)$$

- Consider data matrix X?
- What will be the mean?
- What will be mean of X – mu?
- Y = $a^T X$, mean of Y? var of Y?

(Pdf) $\sim$ $\prod$

Gaus. $\sim$ $p(x \mid \mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

Scalar. MLE $\prod_{i=1}^{N} p(x_i \mid \theta) \rightarrow \theta_{MLE}$

θ_{MLE} Gaussian μ_{MLE}

$$p(x;\mu;\sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{1}{2\sigma^2}(x-\mu_{ML})^2\right\}}}$$

→ am I calc. pdf. of rv. x

→ Am I calc. pdf of N iids. they are realization of $p(x;\mu,\sigma^2)$

$$\theta_{MAP} = \frac{\sum_i x_i + \alpha - 1}{N + \alpha + \beta - 2}$$

$$= \frac{1 + \alpha - 1}{4 + \alpha + \beta - 2}$$

	d	
x_1	1	$x_1 = 1$
x_2	0	$x_2 = 0$
x_3	0	$x_3 = 1$
x_4	0	$x_4 = 1$
	ω_1	ω_2

$\alpha = \beta = 5$

$$\theta_{MAP} = \frac{5}{12}$$

$x_0 = 1$

ω_1 $\omega_2 \longrightarrow \theta_{MAP} = \frac{1}{3}$

x N

$$p(\omega_1 | x) = \frac{p(x | \theta)\, p(\omega_1)}{x}$$

$$= \theta_{MAP}^{x} (1 - \theta_{MAP})^{1-x}\, p(\omega_1)$$

Data matrix $X \in R^{d \times n}$

$$
\begin{array}{c|cccc|c}
 & x_1 & x_2 & x_3 & x_4 & \\
\hline
d_1 & 1 & 0 & 0 & 0 & \frac{1}{4} \\
d_2 & 1 & 0 & 1 & 0 & 2/4 \\
d_3 & 1 & 1 & 1 & 0 & \cancel{\ } 3/4
\end{array}
$$

μ_X

$X - \mu_X$

$$
\begin{array}{ccccc}
3/4 & -1/4 & -1/4 & -1/4 & 0 \\
1/2 & & & & 0 \\
1/4 & & & & 0
\end{array}
$$

$$X \in \mathbb{R}^{d \times n} \qquad X = \begin{array}{cccc} x_i & & & \\ \left[\begin{matrix} x_{11} & \cancel{x_{12}}\; x_{21} & x_{31} & x_{41} \\ x_{12} & x_{22} & x_{32} & x_{42} \end{matrix}\right]_{2 \times 4} \\ x_1 \quad\quad x_2 \quad\quad x_3 \quad\quad x_4 \end{array}$$

$$\mu_X = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad Y = \cancel{a^T x}\; a^T X \qquad a \in \mathbb{R}^{2 \times 1}$$

$$\mu_Y = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{41} \\ x_{12} & & x_{42} \end{bmatrix}$$

$$= a_1 \mu_1 + a_2 \mu_2$$

$$= a^T \mu_X$$

Biased Estimate

$a \in \mathbb{R}^{d\times 1}$

$X = \in \mathbb{R}^{d\times n}$ $\quad Y \in \mathbb{R}^{1\times n}$

*) $Y = a^T X$

$$var(Y) = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \mu_y\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(a^T x_i - a^T \mu_x\right)^2$$

$$= \frac{1}{n-1}\sum_i \underbrace{a^T (x_i - \mu_x)}\,(x_i - \mu_x)^T a$$

$$= a^T\left[\frac{1}{n-1}\sum_i (x_i - \mu_x)(x_i - \mu_x)^T\right] a = a^T S_x a$$

ML estimate of mean is unbiased

$E(x^2) = \mathrm{var}(x) + (E(x))^2$

*) is $E(\sigma^2_{ML}) = \sigma^2$?

$$E\left(\frac{1}{N}\sum_i (x_i - \mu_{ML})^2\right)$$

$$\frac{1}{N}E\left(\sum_i x_i^2 - 2x_i\mu_{ML} + \mu_{ML}^2\right)$$

$$= \frac{1}{N}E\left(\sum_i x_i^2 - 2\mu_{ML}\sum_i x_i + \sum_i \mu_{ML}^2\right)$$

$$= \frac{1}{N}E\left(\sum_i x_i^2 - 2\mu_{ML}^2 N + N\mu_{ML}^2\right)$$

$= \cancel{}$

$$\frac{1}{N}E\left(\sum_i x_i^2\right) - E(\mu_{ML}^2)$$

$$= \frac{1}{N}\sum_i E(x_i^2) - \left\{\mathrm{var}(\mu_{ML}) + \mu^2\right\}$$

$$= \frac{1}{N}\sum_i(\sigma^2 + \mu^2) \cancel{- \mathrm{var}(\mu)} - \mathrm{var}(\mu_{ML}) - \mu^2$$

$$= \sigma^2 + \mu^2 - \mathrm{var}(\mu_{ML}) - \mu^2$$

$$= \sigma^2 - \mathrm{var}(\mu_{ML})$$

$$= \sigma^2 - var\left(\frac{1}{N}\sum_i x_i\right)$$

$$= \sigma^2 - \frac{1}{N^2} var\left(\sum_i x_i\right)$$

$$= \sigma^2 - \frac{1}{N^2} \sum_i var(x_i)$$

$$= \sigma^2 - \frac{1}{N}\sigma^2$$

$$= \sigma^2\left(\frac{N-1}{N}\right) \neq \sigma^2$$

σ^2_{MLE} is biased estimate.

$var(a+b+c) =$

$var(a) + var(b) + var(c)$

if a, b, c are ind.

$$\sigma^2_{MLE} = \frac{1}{N-1}\sum_i (x_i - \mu_{ML})^2$$

$$E(\sigma^2_{MLE}) = \sigma^2$$

- Bias

  - ML estimate for σ^2 is biased

$$E\left[\frac{1}{n}\Sigma(x_i - \bar{x})^2\right] = \frac{n-1}{n}.\sigma^2 \neq \sigma^2$$

  - An elementary unbiased estimator for Σ is:

$$\underbrace{C = \frac{1}{n-1}\sum_{k=1}^{k=n}(x_k - \hat{\mu})(x_k - \hat{\mu})^t}_{\textit{Sample}\ \text{covariance matrix}}$$

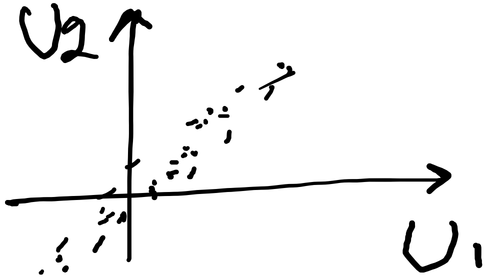
PCA

→ Principal Component Analysis.

- Start with some data matrix, if you are using a discriminant based on MVG, what all things do you need?
- How many multiplications are needed?
- Can you reduce this complexity?

$O(Nd^2)$

- Goal: Project the data onto ~~X~~ orthogonal space such that the projected points have max. variance

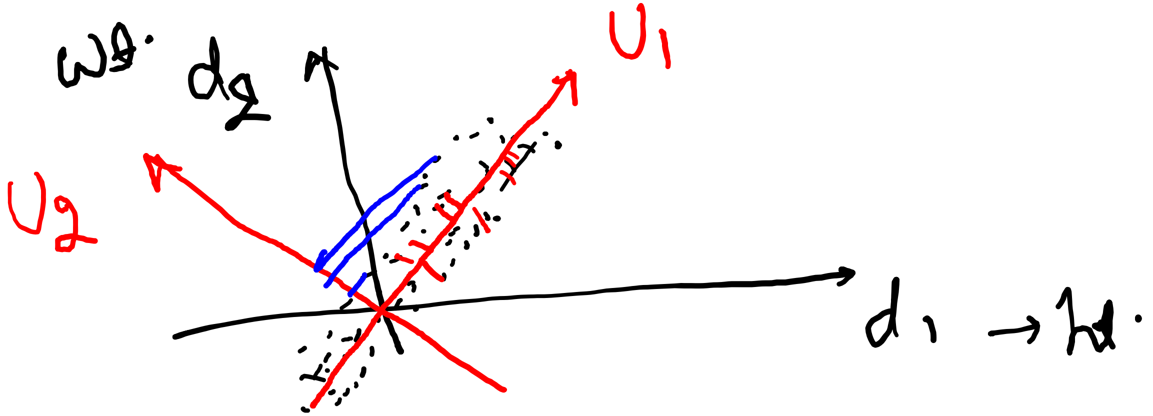


Consider data matrix X and its centralized version obtained by removing the mean.

$X \in \mathbb{R}^{d \times n}$ $\quad \mu_X$

$X \leftarrow X - \mu_X$

$$X \in \mathbb{R}^{2 \times n} \quad \begin{array}{cccc} x_i & x_i & & x_i \\ \left[\begin{array}{cccc} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & & x_{n2} \end{array}\right] & & & \begin{array}{l} \to d_1 \\ \to d_2 \end{array} \end{array}$$



$$Y_i = u_1^T x_i$$

$$y = u_1^T X$$

$$\mu_y = u_1^T \mu_x = 0$$

$$var(Y) = u_1^T S_x u_1$$

$$S_x \to \frac{XX^T}{n-1}$$

$$\max_{u_1} \; u_1^T S_x u_1 \quad s.t. \; u_1^T u_1 = 1$$

$$\max_{u_1} \; u_1^T S_x u_1$$

$$2 S_x u_1 = 0$$

$$u_1 = 0$$