Tutorial 1

Monday, March 25, 2019 5:01 PM

1.3 $u \triangleq \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ $X = a + Qu$ $a = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $Q = \begin{pmatrix} 2 & 1 \\ 2 & 0 \end{pmatrix}$ u_i i.i.d $u_i \sim N(0,1)$.

1.3.a) $\mu_X = E[X]$

$C_X = E[(X-\mu_X)(X-\mu_X)^T]$

$$E[X] = a + QE[u] = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + Q\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \mu_X = a.$$

$$C_X = E[(Qu)u^TQ^T] = Q\underbrace{E[uu^T]}_{I_{2\times2}}Q^T = QQ^T = \begin{pmatrix} 2 & 1 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 4 & 4 \end{pmatrix}$$

1.3.b) ×

1.3.c) $X = \mu_X + P\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ let $C_X = V\Lambda V^T$ $VV^T = I$. $\Lambda \triangleq \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$

then $P = V$ and $z_1 \sim N(0, \lambda_1)$

$z_2 \sim N(0, \lambda_2)$

X is gaussian because z_1, z_2 are Gaussian. Gaussian is fully specified by mean and covariance, hence if the mean and covariance match they follow the same distribution.

$E[X] = \mu_X$ ✓

$$E[(X-\mu_X)(X-\mu_X)^T] = V\,E\left[\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}(z_1\ z_2)\right]V^T = V\Lambda V^T = C_X \checkmark$$

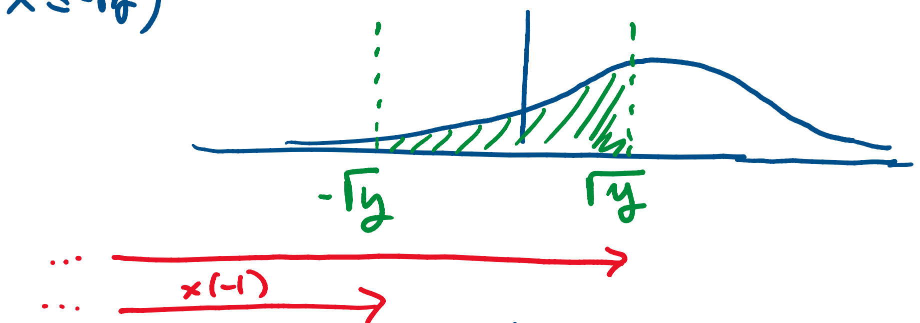
1.3.d) ×

1.8 $X \sim N(\mu, \sigma^2)$ density of Y: $F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) =$

$Y = X^2$ $= P(X \le \sqrt{y}) - P(X \le -\sqrt{y})$

$Y = X^2$

$$= P(X \le \sqrt{y}) - P(X \le -\sqrt{y})$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$



$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \leftarrow \text{valid in solution}$$

when $y = 0 \Rightarrow f_X(\sqrt{y}) = f_X(-\sqrt{y}) \Rightarrow f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi}\sqrt{y}\,\sigma} e^{-y/2\sigma^2}$ χ^2 1 degree of freedom $\underline{\sigma = 1}$.

2.1) Monty Hall Problem:

1 2 3

WLOG assume that we select door 1 (selecting doors 2 or 3 yield the identical argument).

$P(S=1) = P(S=2) = P(S=3) = \frac{1}{3}$. S state of nature.

Then if we select 1; assume the host selects 2, the argument is identical if she opens 3:

$\hookrightarrow P(H=2) = P(H=3) = 1/2$.

$P(H=2 \mid S=1) = P(H=3 \mid S=1) = 1/2$.

$P(H=2 \mid S=2) = 0$

$P(H=2 \mid S=3) = 1$

Then $$P(S=1 \mid H=2) = \frac{P(S=1)P(H=2 \mid S=1)}{P(H=2)} = \frac{1/3 \cdot 1/2}{1/2} = 1/3$$

$P(S=2 \mid H=2) = 0$.

$$P(S=3 \mid H=2) = \frac{P(S=3)P(H=2 \mid S=3)}{P(H=2)} = \frac{1/3 \cdot 1}{1/2} = \boxed{2/3}$$

$$P(S=3|H=2) = \frac{P(S=3)\,P(H=2|S=3)}{P(H=2)} = \frac{1/3}{1/2} = 2/3$$

Strategy is to always change door.

2.5

$\{S_t\}_{t\geq 0}$ First order markov chain $\Rightarrow P(S_t | S_{t-1}, S_{t-2} \dots) = P(S_t | S_{t-1})$.

$$A = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{pmatrix} \qquad a_{ij} = P(S_t = j | S_{t-1} = i)$$

2.5.a) $P_{S_{13}|S_{12}}(k|2)$

$$A = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$

2.5.b) $P_{S_{13}|S_{11},S_{12}}(k|1,2) = P_{S_{13}|S_{12}}(k|2)$

$$A = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$ from 2

2.5.c) $P_{S_{11}}(i) = \frac{1}{3} \quad i = 1,2,3.$

$$P_{S_{11}|S_{12}}(k|2) = \begin{cases} 0.1/0.9 & k=1 \\ 0.8/0.9 & k=2 \end{cases}$$

$$A = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{pmatrix}$$ lead to 2

2.5.d) $P_{S_{11}|S_{10},S_{12}}(k|3,2)$

S_{12} ended up in 2

$\Rightarrow S_{11}$ could not be 3

$$A = \begin{pmatrix} 0.7 & \boxed{0.1} & 0.2 \\ 0.1 & 0.8 & 0.1 \\ \boxed{0.1} & 0 & 0.9 \end{pmatrix}$$

S_{10} was 3 $\Rightarrow S_{11}$ could not be 2

$$P_{S_{11}|S_{10},S_{12}}(k|3,2) = \begin{cases} 1 & k=1 \\ 0 & \text{elsewhere.} \end{cases}$$

'General' way: $P_{S_{11}|S_{10},S_{12}}(k|3,2) = \dfrac{P_{S_{11},S_{12}|S_{10}}}{P_{S_{12}|S_{10}}} = \dfrac{P_{S_{11},S_{12}|S_{10}}}{\sum_{S_{11}=k} P_{S_{11},S_{12}|S_{10}}} =$

- $P_{S_{11},S_{12}|S_{10}} = P_{S_{12}|S_{11},S_{10}} P_{S_{11}|S_{10}} =$

$= P_{S_{12}|S_{11}} P_{S_{11}|S_{10}} = P(S_{12}=2|S_{11}=k) \cdot P_{S_{11}}(k|S_{10}=3)$

$$P_{S_{11},S_{12}|S_{10}}(k,2|3) = P(S_{12}=2|S_{11}=k)\, P_{S_{11}}(k|S_{10}=3) = \begin{cases} 0.1 \cdot 0.9 & k=1 \\ 0.8 \cdot 0 & k=2 \\ 0 \cdot 0.9 & k=3 \end{cases}$$

3.3 $S \in \{1,\dots,N\}$ known: $P_S(j)$ priors

$f_{X|S}(x|j)$ likelihood function

Design classifier $d(x): \mathcal{X} \to \{1,\dots,N,N+1\}$ class $N+1$ meaning 'cannot classify'.
that minimizes expected loss given loss function:

$$L(d(x)=i|S=j) = \begin{cases} 0, & i=j=1,\dots,N \\ r, & i=N+1, j=1,\dots,N \\ c, & \text{otherwise} \end{cases}$$

3.3.a.

$d(x) := \arg\min_i R[i|x]$ N

$$d(x) := \arg\min_i R(i|x)$$

with $R(i|x) = \begin{cases} \sum_{j=1}^{N} L(d(x)=i \mid S=j)\, P_{S|X}(j|x), & i=1,\dots,N \\ r, & i=N+1 \end{cases}$

Hence whenever $\sum_{j=1}^{N} L(d(x)=i \mid S=j)\, P_{S|X}(j|x) \leq r$ for at least one $i=1,\dots,N$ we will minimize LHS otherwise we will choose r.

$$0 \cdot P_{S|X}(i|x) + \sum_{j \neq i} c\, P_{S|X}(j|x) = c\,(1 - P_{S|X}(i|x)) \leq r \qquad 1 - P_{S|X}(i|x) \leq \frac{r}{c}$$

$1 - \frac{r}{c} \leq P_{S|X}(i|x)$.

Hence if for at least one i $1 - r/c \leq P_{S|X}(i|x)$ we select i s.t. $i = \arg\min_i \sum_{j=1}^{N} L(d(x)=i \mid S=j)\, P_{S|X}(j|x)$

$= \arg\min_x \sum_{j \neq i} c\, P_{S|X}(j|x) = \cancel{c}\,(1 - P_{S|X}(i|x)) = \arg\max_i P_{S|X}(i|x)$.

Hence,

$$d(x) = \begin{cases} \arg\max_i P_{S|X}(i|x) & \text{if } P_{S|X}(i|x) \geq 1 - r/c \\ N+1 & \text{otherwise.} \end{cases}$$

(if holds for max, holds for at least 1).

3.3.b What happens if $r=0$ or $c \to \infty$?

$\exists$ no reason to classify as it will result in an average higher cost. Then the classifier will

$\exists$ no reason to clarify as it will result in an average higher cost. Then the classifier will prefer to never pick anything.

3.3.c If $r > c$?

The do not clarify option is more expensive than making a mistake and therefore we will always clarify.

3.3.d. Show that $g_i(x) = f_{X|S}(x|i)P_S(i) \quad i = 1,\dots,N$

$$g_{N+1}(x) = \left(1 - \frac{r}{c}\right)\sum_{j=1}^{N} g_j(x)$$

give the minimum cost.

For this to be true it needs to happen that (a) $P_{S|X}(s=i|x) \propto g_i(x)$

and that (b) $g_{N+1}(x) \geq g_i(x) \ \forall i = 1,\dots,N$ when $P_{S|X}(i|x) \leq 1 - r/c \quad \forall i = 1,\dots,N$.

(a)

We first establish that $P_{S|X}(s=i|x) \propto g_i(x)$

$$g_i(x) = f_{X|S}(x|i)P_S(i) = P_{S|X}(s=i|x)f_X(x) \Rightarrow g_i(x) \propto P_{S|X}(s=i|x).$$

(b) $$\sum_{i=1}^{N} g_i(x) = \sum_{i=1}^{N} f_{X|S}(x|i)P_S(i) = f_X(x)$$

Then if $$g_{N+1}(x) = \left(1-\frac{r}{c}\right)\sum_{j=1}^{N} g_j(x) = \left(1-\frac{r}{c}\right)\cancel{f_X(x)} \geq P_{S|X}(s=i|x)\cancel{f_X(x)} \quad \forall i$$

condition (b) is also fulfilled.

3.3.e. Sketch discriminants when $N=2$:

$S=1$	$\mathcal{N}(1,1)$
$S=2$	$\mathcal{N}(-1,1)$

$P(S=1) = P(S=2) = 1/2$

$S=2 \mid \mathcal{N}(-1,1)$

choose 'not classify' when $P_{S|X}(i|x) \leq 3/4$.

$$P_{S|X}(1|x) = \frac{f_{X|S}(x|1)\,P(S=1)}{f_X(x)}$$

Abuse of notation $\mathcal{N}(\mu, C)(x)$

$$P_{S|X}(2|x) = \frac{f_{X|S}(x|2)\,P(S=2)}{f_X(x)}$$

$$P_{S|X}(1|x) = \frac{\frac{1}{2}\mathcal{N}(1,1)(x)}{\frac{1}{2}\mathcal{N}(1,1)(x) + \frac{1}{2}\mathcal{N}(1,-1)(x)}$$

$$P_{S|X}(2|x) = \frac{\frac{1}{2}\mathcal{N}(-1,1)(x)}{\frac{1}{2}\mathcal{N}(1,1)(x) + \frac{1}{2}\mathcal{N}(-1,1)(x)}$$

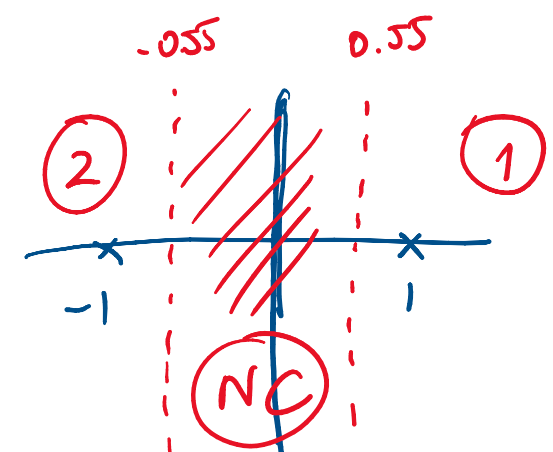
$$P_{S|X}(1|x) = \frac{\mathcal{N}(1,1)(x)}{\mathcal{N}(1,1)(x) + \mathcal{N}(1,-1)(x)} = 3/4 \Rightarrow \mathcal{N}(1,1)(x) = 3\,\mathcal{N}(1,-1)(x).$$

$$\cancel{\frac{1}{\sqrt{2\pi}\,(1)}}\, e^{-(x-1)^2/2} = \cancel{\frac{1}{\sqrt{2\pi}\,(1)}}\, e^{-(x+1)^2/2} \cdot e^{\log(3)}$$

$$\Rightarrow (x-1)^2/2 = (x+1)^2/2 - \log(3)$$

$$\cancel{x^2} - 2x + \cancel{1} = \cancel{x^2} + 2x + \cancel{1} - 2\log(3)$$

$$4x = 2\log(3) \qquad x = \frac{1}{2}\log(3) \approx 0.55$$



3.3f | Probability of "reject" decision in the situation 3.3 e ?

3.3.f. Probability of "reject" decision in the situation 3.3 e?

$$P(\text{reject}) = \int_{-\frac{1}{2}\log(3)}^{\frac{1}{2}\log(3)} f_X(x)\,dx = \frac{1}{2}\int_{-\frac{1}{2}\log(3)}^{\frac{1}{2}\log(3)} \left(\mathcal{N}(1,1)(x) + \mathcal{N}(-1,1)\right)dx$$

$$= \frac{1}{2}\int_{-\frac{1}{2}\log(3)-1}^{\frac{1}{2}\log(3)-1} \mathcal{N}(0,1)(x)\,dx + \frac{1}{2}\int_{-\frac{1}{2}\log(3)+1}^{\frac{1}{2}\log(3)+1} \mathcal{N}(0,1)(x)\,dx$$

$$= \frac{1}{2}\int_{-\frac{1}{2}\log(3)-1}^{\infty} \mathcal{N}(0,1)(x)\,dx - \int_{\frac{1}{2}\log(3)-1}^{\infty} \mathcal{N}(0,1)(x)\,dx \quad + \text{analogous}$$

$$= \frac{1}{2}\, Q\left(-\left(\tfrac{1}{2}\log(3)+1\right)\right) - \frac{1}{2} Q\left(\tfrac{1}{2}\log(3)-1\right) + \frac{1}{2} Q\left(-\tfrac{1}{2}\log(3)+1\right) - \frac{1}{2} Q\left(\tfrac{1}{2}\log(3)+1\right)$$

$$= \frac{1}{2}\left(\cancel{1} - Q\left(\tfrac{1}{2}\log(3)+1\right)\right) - \frac{1}{2} Q\left(\tfrac{1}{2}\log(3)+1\right) + \frac{1}{2} Q\left(-\tfrac{1}{2}\log(3)+1\right)$$

$$-\frac{1}{2}\left(\cancel{1} - Q\left(1-\tfrac{1}{2}\log(3)\right)\right) = Q\left(1-\tfrac{1}{2}\log(3)\right) - Q\left(1+\tfrac{1}{2}\log(3)\right)$$

range of relations r/c will the reject decision never be chosen? $\frac{r}{c} > 0.5$.

Because one of the two $P(S|X)$ will always be equal or larger than 0.5.

3.13 $X = (X_1, \dots, X_T)$ input data

$E[X_t] = \begin{cases} -1 & s=0 \\ \dots \end{cases}$ → equal probabilities

$\operatorname{Var}(X_t) = 4$ for both.

3.13 $X = (x_1, \dots, x_T)$ input data

Generated with either state $S=0$ or $S=1$

$E[X_t] = \begin{cases} -1 & S=0 \\ +1 & S=1 \end{cases}$ Gaussian

$\text{var}(X_t) = 4$ for both.

$T=9$.

3.13.a Smallest probability of error for ideal classifier?

log likelihood functions : $\log \left(\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi} \cdot 2} e^{-(x_t - \mu_i)^2 / 2 \cdot 4} \right)$

only difference between the two classes : exponent $- \sum_{t=1}^{T} (x_t - \mu_i)^2$

Hence if $\sum_{t=1}^{T} (x_t - \mu_1)^2 \geq \sum_{t=1}^{T} (x_t - \mu_0)^2$

choose 0

$\Rightarrow \sum_{t=1}^{T} (\cancel{x_t^2} - 2 x_t \mu_1 + \cancel{\mu_1^2}) \overset{0}{\geq} \sum_{t=1}^{T} (\cancel{x_t^2} - 2 x_t \mu_0 + \cancel{\mu_0^2})$

$\sum_{t=1}^{T} - x_t \overset{0}{\geq} \sum_{t=1}^{T} x_t$ hence if $\boxed{\sum_{t=1}^{T} x_t \leq 0 \text{ we choose } S=0}$

x_i are iid and Gaussian $\Rightarrow \sum_{t=1}^{T} x_t$ i.i.d. and Gaussian.

$Y = \sum_{t=1}^{T} x_t$ $\quad E[Y|S=0] = -T$ $\quad \text{var}[Y] = T4$

$$Y \triangleq \sum_{t=1}^{T} X_t = \begin{cases} E[Y|S=0] = -T \\ E[Y|S=1] = T \end{cases} \qquad \mathrm{var}[Y] = T4$$

$Y \leq 0$ choose $S=0$.

$$P(\text{error}) = P(Y \geq 0 \mid S=0)\,P(S=0) + P(Y<0 \mid S=1)\,P(S=1)$$
$$= P(Y \geq 0 \mid S=0)$$

$$= \int_0^{\infty} \frac{1}{\sqrt{(2\pi)}\,2\sqrt{T}} e^{-(y+T)^2/2\cdot T4}\, dy = \left\{ z = \frac{y+T}{2\sqrt{T}} \right\} = \int_{\sqrt{T}/2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z)^2/2}\, dz = Q\left(\sqrt{T}/2\right).$$

3.13.6. X_t a discrete 'vote' for $S=1$ if $X_t > 0$
$S=0$ otherwise

$Z = \# X_t > 0$ $\quad S=0$ iff $Z < T/2$
$S=1$ otherwise.

Probability of error. $\quad T=9$

$$P(\text{error}) = P(\underbrace{Z \geq T/2}_{Z \geq 5} \mid S=0)\,P(S=0) + P(\underbrace{Z < T/2}_{Z \leq 4} \mid S=1)\,P(S=1) = P(Z \leq 4 \mid S=1).$$

There are a total of $\binom{T}{n}$ # of ways n objects can be chosen among T objects (disregarding order).

$$P(x_t > 0 \mid S=0) = \int_0^\infty \frac{1}{\sqrt{2\pi}\cdot 2} e^{-(x-1)^2/2\cdot 4} = Q(-\tfrac{1}{2}) = 1 - Q(\tfrac{1}{2}) = 1-\varepsilon$$

$$P(x_t > 0 \mid S=1) = Q(\tfrac{1}{2}) = \varepsilon$$

$$P(z=n \mid S=0) = \binom{T}{n}(1-\varepsilon)^n \varepsilon^{T-n} \qquad P(z=n \mid S=1) = \binom{T}{n}(\varepsilon)^n(1-\varepsilon)^{T-n}$$

$$P(z \geq T/2 \mid S=0) = \sum_{n=\lceil T/2 \rceil}^{T} \binom{T}{n}(1-\varepsilon)^n(\varepsilon)^{T-n}.$$

$$P(z < T/2 \mid S=1) = \sum_{n=1}^{\lceil T/2 \rceil - 1} \binom{T}{n}(\varepsilon)^n(\varepsilon)^{T-n}$$

$$P(z \leq 4 \mid S=1) = \sum_{n=1}^{4} \binom{9}{n}(\varepsilon)^n(\varepsilon)^{9-n} \approx \underline{\underline{0.11}}$$

3.6

0 fair

1 $p(\text{head} \mid S=1) = p > \frac{1}{2}$ $\quad p(\text{tail} \mid S=1) = 1-p.$

K tosses $\underline{x}$ k heads

3.6.a. $g(\underline{x}) > 0$ coin 1
$g(\underline{x}) \leq 0$ coin 0.

# of heads in a sequence of length K

$$P(H=k \mid S=0) = (0.5)^k (0.5)^{K-k} = (0.5)^K$$
$$P(H=k \mid S=1) = (p)^k (1-p)^{K-k}$$

$g(\underline{x}) > 0$ coin 1 $\rightarrow$ $P(H=k \mid S=1) > P(H=k \mid S=0)$ log monotonically increasing

$$g(x) = k\ln(p) + (K-k)\ln(1-p) - K\ln(0.5)$$
$$= k\left(\ln(p) - \ln(1-p)\right) + K\left(\ln(1-p) + \ln(2)\right)$$

$$k_{th}\left(\ln(p) - \ln(1-p)\right) + K\left(\ln(1-p) + \ln(2)\right) = 0$$

$$k_{th} = \frac{-K\left(\ln(1-p) + \ln(2)\right)}{-\ln(1-p) + \ln(p)} = K \cdot \frac{\ln\left(\frac{1}{2(1-p)}\right)}{\ln\left(\frac{p}{1-p}\right)}$$

3.6.b. average probability of error for $k=5$ and $p=0.6$.

$k_{th} = 2.75$.

$$P(\text{error}) = P(H \geq 3 \mid p=0.5)\tfrac{1}{2} + P(H \leq 2 \mid p=0.6)\tfrac{1}{2} \approx 0.409.$$

$$P(\text{error}) = P(\underbrace{H \geq 5 \mid p = \ldots}_{S=0})\,\tfrac{1}{2} + P(\underbrace{H < 5 \mid p = \ldots}_{S=1})\,\tfrac{1}{2}$$