ES16BTECH11005
Bhavsar Jeel Alpesh

HW 1

1) Find $\hat{y}(\bar{x})$ such that $E[(y-\hat{y}(\bar{x}))^2]$ is minimizes

$$E[(y-\hat{y}(x))^2] = \iint (y-\hat{y}(x))^2 \, p(\bar{x},\bar{y}) \, dx\,dy$$

$$\frac{d\,E[\;]}{d(\hat{y}(x))} = 0$$

$$\frac{dE}{d\hat{y}(x)} = -2\int (y-\hat{y}(x))\, p(\bar{x},\bar{y})\, dy = 0$$

• From Bayes, $p(\bar{x},\bar{y}) = p(y|\bar{x})\,p(\bar{x})$

$\cancel{\int(y-\hat{y}(x}$

$$p(x)\underbrace{\int y\, p(y|x)\, dy}_{E(y|x)} - p(x)\int \underbrace{\hat{y}(x)}_{\text{does not depend on } y} p(y|x)\,dy = 0$$

$$p(x)\,E(y|x) - p(x)\,\hat{y}(x)\underbrace{\int p(y|x)\,dy}_{=1} = 0$$

So optimal estimator $\boxed{\hat{y}(x) = E(y|x)}$

2) Arbitrary model $\hat{y}(\bar{x})$

Optimal estimator $y^*(\bar{x}) = E(y|\bar{x})$

$y \rightarrow$ ground truth label   from problem I

Cost function $= E(\hat{y}(\bar{x}) - y)^2$

$$E[(\hat{y}(x) - y)^2] = E[(\hat{y}(x) - y^*(\bar{x}) - (y - y^*(x)))^2]$$

$$= E[(\hat{y}(x) - y^*(x))^2] + E[(y - y^*(x))^2]$$

$$- 2E[(\hat{y}(x) - y^*(x))(y - y^*(x))]$$

Let it be equal to z

$$z = 2\iint (\hat{y}(x) - y^*(x)) \cdot (y^*(x) - y)\, p(\bar{x}, y)\, dx\, dy$$

From Bayes, $p(\bar{x}, y) = p(y|x)\, p(x)$

~~$z = 2\iint(\hat{y}$~~

$$z = 2\int (\hat{y}(x) - y^*(x)) \cdot \int (y^*(x) - y)\, p(y|x)\, dy\; p(x)\, dx$$

$$= 2\int (\hat{y}(x) - y^*(x)) \cdot \left[\int y^*(x)\, p(y|x)\, dy - \int y\, p(y|x)\, dy\right] p(x)\, dx$$

$$y^*(x) = E[y|x]$$

$$\int y^*(x)\, p(y|x)\, dy = E\big[E[y|x]\big]$$
$$= E[y|x]$$

$$\int y\, p(y|x)\, dy = E[y|x]$$

So $$z = 2\int (\hat{y}(x) - y^*(x)) \cdot \big[E[y|x] - E[y|x]\big]\, p(x)\, dx$$

$$z = 0$$

So $$E[(\hat{y}(x) - y)^2] = E[(\hat{y}(x) - y^*(x))^2] + E[(y - y^*(x))^2]$$

Add and subtract $E(\hat{y}(x))$

$$E[(\hat{y}(x) - y^*(x))^2] = E\Big[\big((\hat{y}(x) - E(\hat{y}(x))) - (y^*(x) - E(\hat{y}(x)))\big)^2\Big]$$

$$= E[(\hat{y}(x) - E(\hat{y}(x)))^2] + E[(y^*(x) - E(\hat{y}(x)))^2]$$
$$-2E\big[(\hat{y}(x) - E(\hat{y}(x))) \cdot (y^*(x) - E(\hat{y}(x)))\big]$$

This is similar to how we showed $E(y^*(x) - y) = 0$ when $y^*(x) = E[y|x]$

So $$\boxed{-2E\big[(\hat{y}(x) - E(\hat{y}(x))) \cdot (y^*(x) - E(\hat{y}(x)))\big] = 0}$$

Hence

$$E\left[(\hat{y}(x) - y^*(x))^2\right] = E\left[(\hat{y}(x) - E(\hat{y}(x)))^2\right] + E\left[(y^*(x) - E(\hat{y}(x)))^2\right]$$

So

$$E\left[(\hat{y}(x) - y)^2\right] = \underbrace{E\left[(\hat{y}(x) - E(\hat{y}(x)))^2\right]}_{\text{variance}} + \underbrace{E\left[(y^*(x) - E(\hat{y}(x)))^2\right]}_{\text{bias}} + \underbrace{E\left[(y - y^*(x))^2\right]}_{\text{noise}}$$

k^{th} column → belongs to C_k

3) $$Y = \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & \cdots 1 & \cdots & \cdots & 0 \\ \vdots & & & & & \\ \vdots & & & & & \end{bmatrix}_{N \times K}$$

j^{th} column → belongs to C_j

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1D} \\ 1 & & & \\ \vdots & & & \\ 1 & x_{N1} & \cdots & x_{ND} \end{bmatrix}_{N \times (D+1)}$$

$$W = \begin{bmatrix} w_{01} & w_{02} & \cdots & w_{0K} \\ w_{11} & & & \\ \vdots & & & \\ w_{D1} & w_{D2} & \cdots & w_{DK} \end{bmatrix}_{(D+1) \times K}$$

$$\hat{Y} = XW$$

Cost function $= E(\vec{w})$

$$E(w) = (Y - \hat{Y})^2$$

$$E(w) = (Y - XW)^T (Y - XW)$$

$$\nabla E(w) = 0$$

From Linear Regression solution,

$$W = (X^T X)^{-1} X^T Y$$

4) Fischer's linear discriminant for 2 class classifier case.

Take the D dimensional input vector X and project it to 1D

$$y = W^T x$$

Say N_1 points belong to Class C_1
N_2 " " " " C_2

The corresponding mean vectors:

$$\bar{m}_1 = \frac{1}{N_1} \sum_{i \in C_1} x_i \qquad \bar{m}_2 = \frac{1}{N_2} \sum_{i \in C_2} x_i$$

$$m_2 - m_1 = W^T(\bar{m}_2 - \bar{m}_1)$$

i.e. $m_i = W^T \bar{m}_i$

The within class variance is

$$S_k^2 = \sum_{i \in C_k} (y_i - m_k)^2$$

$$y_i = W^T x_i$$

Total within class variance $= S_1^2 + S_2^2$

Fisher Criterion is ratio of between class variance to within class variance

$$J(w) = \frac{(m_2 - m_1)^2}{S_1^2 + S_2^2}$$

$$J(w) = \frac{w^T (m_2 - m_1)(m_2 - m_1)^T w}{w^T \left(\sum_{i \in C_1} (x_i - \bar{m}_1)(x_i - \bar{m}_1)^T + \sum_{i \in C_2} (x_i - \bar{m}_2)(\bar{x}_i - \bar{m}_2)^T \right) w}$$

Maximize $J(w)$

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

Differentiating w.r.t w and equating it to 0,

$$\frac{(w^T S_B w) S_B w - (w^T S_B w) S_w w}{(w^T S_w w)^T (w^T S_w w)} = 0$$

$$(w^T S_B w) S_w w = (w^T S_w w) S_B w$$

$S_B w$ ||^{el} to $(\bar{m}_2 - \bar{m}_1)$

$$\boxed{w = S_w^{-1} (\bar{m}_2 - \bar{m}_1)}$$

optimum w that maximizes $J(w)$

Scanned by CamScanner

5) ~~With the given loss function,~~ we can ~~write define $y^*(x)$~~

5) $y^*(x) = \underset{\hat{y}(x)}{\text{Arg min}}\ E\left[L(y, \hat{y}(x))\right]$

$$L(y, \hat{y}(x)) = \begin{cases} 0 & y = \hat{y}(x) \\ 1 & y \neq \hat{y}(x) \end{cases}$$

$$E_{xy}\left[L(y, \hat{y}(x))\right] = E_x\left[\sum_{y \in C_k} L(y, \hat{y}(x))\, p(y=k|x)\right]$$

$$\because E_{xy}\left[f(x,y)\right] = E_x\left[E_{y|x}\left[f(x,y)\right]\right]$$

So $y^*(x) = \underset{\hat{y}(x)}{\text{Arg min}}\ E_x\left[\sum_{y \in C_k} L(y, \hat{y}(x))\, p(y=k|x)\right]$

$$= \underset{\hat{y}(x)}{\text{Arg min}}\ E_x\Big[L(y=1, \hat{y}(x))\, p(y=1|x) + L(y=2, \hat{y}(x))\, p(y=2|x) + \cdots + L(y=k, \hat{y}(x))\, p(y=k|x)\Big]$$

From the def of $L(y, \hat{y}(x))$

If $\hat{y}(x) = y = j$, ~~the rest of the terms will be 0~~. $L(y, \hat{y}(x))$ will be 1 for the rest of the terms and 0 for $y = j$

$$\cancel{y^*(x) = \underset{\hat{y}(x)}{\mathrm{Argmin}}\, E_x\left[0 + 0 \ldots + p(y=j|x) + 0 \ldots + 0\right]}$$

$$= A$$

$$\cancel{y^*(x) = \underset{\hat{y}(x)}{\mathrm{Argmin}}\, E_x\left[p(y \sum\right.}$$

$$\cancel{y^*(x) = \underset{\hat{y}(x)}{\mathrm{Argmin}}\, E_x\left[\sum_{y \in C_k}\right.}$$

$$y^*(x) = \underset{\hat{y}(x)}{\mathrm{Argmin}}\, E_x\left[1 - p(y=j|x)\right]$$

$$\boxed{y^*(x) = \underset{y \in C_k}{\mathrm{Argmax}}\, p(y=k|\bar{x})}$$