॥ भगवान् श्री रामकृष्ण शरणम् ॥

Tutorial 02 : 06/02/23

Agenda : Bayes Classifier
MinMax classifier
NP Classifier

Bayes Classifier:

M-Classes
arbitrary loss func

$y(x) \in \{C_0, \cdots, C_{M-1}\}$ m class labels.

$h(x) \in \{\alpha_0, \cdots, \alpha_{K-1}\}$ output of Classifier.

$L(\alpha_j, C_k)$: Loss when classifier says α_j
true class in C_k.

$$R(h) = \mathbb{E}_{XY} \, L(h(x), y(x))$$

we want a classifier with least risk value.

$R(\alpha_i|x)$: Expected loss $h(x) = \alpha_i$, cond^n on x

$$R(h) = \mathbb{E}_{XY}[L(h(x), y(x))]$$

$$= \mathbb{E}_{X} \underbrace{\mathbb{E}_{Y|X}[L(h(x), y(x)) \,|\, x]}_{R(h(x)|x)}$$

$$R(\alpha_i|x) = \mathop{E}_{Y|X}\left[L(h(x), y(x)) \,|\, h(x) = \alpha_i, x\right]$$

$$= \mathop{E}_{Y|X}\left[L(\alpha_i, y(x)) \,/ x\right]$$

$$R(\alpha_i|x) = \sum_{j=0}^{M-1} L(\alpha_i, c_j)\, P[y(x) = c_j \,/ x]$$

for a general case.

$$R(h(x)/x) = \sum_{j=0}^{M-1} L(h(x), c_j)\, P[y(x) = c_j \,/ x]$$

$$R(h) = \int R(h(x)/x) \underbrace{f(x)}_{\text{density of } X} dx$$

* Optimal classifier

$$R(h(x)/x) \leq R(h'(x)/x) \quad \forall\, h'$$

* h_B : Bayes classifier.

$$h_B(x) = \alpha_j$$

$$\sum_{k=0}^{M-1} L(\alpha_j, c_k)\, P[y(x) = c_k \,/ x] \leq$$

$$\sum_{k=0}^{M-1} L(\alpha_t, C_k)\, P[y(x) = C_k \mid x] \quad \forall t$$

$$* \quad R(h_B(x) \mid x) \leq R(h(x) \mid x)$$

2 class case.

$$L(\alpha_0, C_0)\, P[y(x)=0 \mid x] + L(\alpha_0, C_1)\, P[y(x)=1 \mid x]$$

$$\leq L(\alpha_1, C_0)\, P[y(x)=0 \mid x] + L(\alpha_1, C_1)\, P[y(x)=1 \mid x]$$

Consider 0-1 Loss.

$$\frac{P[y(x)=0 \mid x]}{P[y(x)=1 \mid x]} \geq \frac{L(\alpha_0, C_1)}{L(\alpha_1, C_0)}$$

$$\underbrace{\frac{P[x \mid y(x)=0]}{P[x \mid y(x)=1]}}_{\text{Likelihood Ration}} \; \underbrace{\frac{P[y(x)=0]}{P[y(x)=1]}}_{\text{Prior Ration.}} \geq \beta$$

M- Class Case with 0-1 Loss

$$R(\alpha_i | x) = \sum_{j=0}^{M-1} L(\alpha_i, C_j)\, P[y(x) = C_j | x]$$

Considering 0-1 Loss

$$= \sum_{j \neq i} P[y(x) = C_j / x]$$

$$= 1 - P[y(x) = C_i | x]$$

Consider $h_B(x) = \alpha_i$

$$\{1 - P[y(x) = C_i | x]\} \leq \{1 - P[y(x) = j | x]\}$$

$$P[y(x) = C_i | x] \geq P[y(x) = C_j | x] \quad \forall j$$

$p_i(x) \Rightarrow P[x / y(x) = C_i]$

$p(y = i) \Rightarrow$ Prior.

$$p_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ - \frac{(x - \mu_i)^2}{2\sigma_i^2} \right\} \qquad i = 0, 1$$

$h_B(x) = 0$ if

$p(y=0)\, p_0(x)\, L(1,0) > p(y=1)\, p_1(x)\, L(0,1)$

taking log.

$\ln\{p(y=0)\, L(1,0)\} + \ln p_0(x) >$

$\ln\{p(y=1)\, L(0,1)\} + \ln p_1(x)$

$$\ln\{p(y=0)\, L(1,0)\} - \ln(\sigma_0) - \frac{1}{2}\ln 2\pi - \frac{(x-\mu_0)^2}{2\sigma_0^2} >$$

$$\ln\{p(y=1)\, L(0,1)\} - \ln(\sigma_1) - \frac{1}{2}\ln 2\pi - \frac{(x-\mu_1)^2}{2\sigma_1^2}$$

$$\frac{1}{2}x^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) + x\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)$$

$\mu_0 = \mu_1 = 0$

$$+ \frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2}\right) + \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \ln\left\{\frac{p(y=0)\, L(1,0)}{p(y=1)\, L(0,1)}\right\}$$

$$> 0$$

$\sigma_0 = \sigma_1 = \sigma,\quad p(y=0) = p(y=1)$

$L(1,0) = L(0,1)$

$$\frac{x}{\sigma^2}(\mu_0 - \mu_1) - \frac{1}{2\sigma^2}(\mu_0^2 - \mu_1^2) > 0$$

$x > \frac{\mu_0 + \mu_1}{2}$ when $\mu_0 > \mu_1$

* Bayes classifier : - Prior
Class Cond^n

Minimize the maximum possible overall Risk.

"Min Max Classifier."

2 class problem :

$$R(h) = \int_{R_1(h)} L(1,0)\, P[y=0]\, \underbrace{p_0(x)}_{\to P[x|y=0]}\, dx \; +$$

$$\int_{R_0(h)} L(0,1)\, P[y=1]\, p_1(x)\, dx.$$

wrt $P[y=0] = 1 - P[y=1]$

$$R(h) = L(1,0)\, P[y=0] \int_{R_1(h)} p_0(x)\, dx \; +$$

$$L(0,1)\,(1 - P[y=0]) \int_{R_0(h)} p_1(x)\, dx.$$

$$= L(0,1) \int_{R_0(h)} p_1(x)\, dx \; +$$

$$P[y=0] \left\{ L(1,0) \int_{R_1} p_0(x)\, dx - L(0,1) \int_{R_0} p_1(x)\, dx \right\}$$

Consider a Classifier s.t.

$$L(1,0) \int_{R_1(h)} p_0(x)\, dx = L(0,1) \int_{R_0(h)} p_1(x)\, dx.$$

$$\int_{R_1(h)} p_0(x)\, dx = \int_{R_0(h)} p_1(x)\, dx \quad \Big/ \; L(1,0) = L(0,1)$$

$$p_i(x) \sim N(x; \mu_i, \sigma_i)$$

$$h(x) = 0 \quad \text{iff} \quad x < a$$

fix a threshold, to satisfy minimax criteria.

$$\int_a^{\infty} p_0(x)\, dx = \int_{-\infty}^{a} p_1(x)\, dx.$$

$$z = \frac{x - \mu_0}{\sigma_0} \quad \text{for LHS.}$$

$$z = \frac{x - \mu_1}{\sigma_1} \quad \text{for RHS.}$$

// with the following change of variable.

$$1 - \phi\left(\frac{a - \mu_0}{\sigma_0}\right) = \phi\left(\frac{a - \mu_1}{\sigma_1}\right)$$

$$1 - \phi(x) = \phi(-x)$$

$$\phi\left(\frac{\mu_0 - a}{\sigma_0}\right) = \phi\left(\frac{a - \mu_1}{\sigma_1}\right)$$

$$\frac{\mu_0 - a}{\sigma_0} = \frac{a - \mu_0}{\sigma_1} \Rightarrow a = \frac{\mu_0 \sigma_1 + \mu_1 \sigma_0}{\sigma_0 + \sigma_1}$$

Note:

* when σ_1 & σ_0 are not same Bayes classifier was quadratic. But minimax classifier is linear.

Tutorial 2 : 15/2/2023

Neyman Pearson criteria:

	+ve	-ve	← Actual
+ve	TP	False +ve	
-ve	False -ve	TN	
↑ Predicted			

Type 1 error : False +ve
Type 2 error : False -ve.

minimize Type2 error under a constraint that Type1 error is below some threshold. : NP criterion

Bound Type 1 error $\alpha \in (0,1)$

h_{NP} is characterized by.

1) $P[h_{NP}=1 \mid x \in C_0] \leq \alpha$

2) $P[h_{NP}=0 \mid x \in C_1] \leq P[h(x)=0 \mid x \in C_1]$
$\forall\ h$ s.t $P[h(x)=1 \mid x \in C_0] \leq \alpha$

$f_i(x) = P[x \mid y=i]$ Class Cond^{n}L
$q_i(x) = P[y=i \mid x]$ Label posterior.

NP Classifier:

$$h_{NP}(x) = \begin{cases} 1 & \text{if } \frac{f_1(x)}{f_0(x)} > K \\ 0 & \text{otherwise} \end{cases}$$

where K is s.t

$$P\left[\frac{f_1(x)}{f_0(x)} \le K \mid x \in C_0\right] = 1-\alpha$$

$$P\left[\frac{f_1(x)}{f_0(x)} > K \mid x \in C_0\right] = \alpha \qquad \text{By Const}^n$$

we need to satisfy the minimum Type 2 Cond^n.

Let h be any classifier s.t

$$P[h(x) = 1 \mid x \in C_0] \le \alpha$$

we need to show

$$P[h_{NP}(x) = 0 \mid x \in C_1] \le P[h(x) = 0 \mid x \in C_1]$$

or we can equivalently say.

$$P[h_{NP}(x) = 1 \mid x \in C_1] \ge P[h(x) = 1 \mid x \in C_1]$$

Consider an Integral "$\mathcal{J}$"

$$\mathcal{J} = \int_{\mathbb{R}^n} \{h_{NP}(x) - h(x)\}\{f_1(x) - k f_0(x)\}\, dx.$$

$$= \int_{f_1 > k f_0} \{h_{NP}(x) - h(x)\}\{f_1(x) - k f_0(x)\}\, dx. +$$

$$\int_{f_1 \leq k f_0} \{h_{NP}(x) - h(x)\}\{f_1(x) - k f_0(x)\}\, dx.$$

$f_1 > k f_0 \; : \Rightarrow \quad \dfrac{f_1}{f_0} > k \Rightarrow h_{NP}(x) = 1$

$h_{NP}(x) - h(x) \geq 0 \Rightarrow$ term 1 ≥ 0

$f_1 < k f_0 \Rightarrow h_{NP}(x) = 0$

$$\left. \begin{aligned} 0 - h(x) &\leq 0 \\ f_1 - K f_0(x) &\leq 0 \end{aligned} \right\} \Rightarrow \text{term } 2 \geq 0$$

$\mathcal{J} \geq 0$

$$\int h_{NP}(x) f_1(x)\, dx - \int h(x) f_1(x) \geq$$

$$k \left\{ \int h_{NP}(x) f_0(x)\, dx - \int h(x) f_0(x) \right\}$$

$h_{NP}, h(x) \in \{0,1\}$

$$\overbrace{\int h_{NP}(x) f_1(x)\,dx}^{P[h_{NP}(x)=1 / x \in C_1]} - \overbrace{\int h(x) f_1(x)}^{P[h(x)=1 / x \in C_1]} \geq$$

$$K \left\{ \int h_{NP}(x) f_0(x)\,dx - \int h(x) f_0(x) \right\}$$

$$P[h_{NP}(x)=1 / x \in C_1] - P[h(x)=1 / x \in C_1] \geq$$

$$\underbrace{K \left\{ \overbrace{P[h_{NP}(x)=1 / x \in C_0]} - \overbrace{P[h(x)=1 / x \in C_0]} \right\}}_{\text{non -ve}}$$

why??

$\Rightarrow P[h_{NP}(x)=1 / x \in C_1] - P[h(x)=1 / x \in C_1] \geq 0$

example: $x \in \mathbb{R}$ $\quad \left.\begin{matrix} f_1 \\ f_0 \end{matrix}\right\} N$ with equal Variance.

$\mu_0 < \mu_1$

$h(x)$:

$X > \tau$: NP classifier

$x \in C_1$

0-1 Loss

equal prior.

$$\boxed{\int_{\tau}^{\infty} f_0(x)\,dx = \underline{\underline{\alpha}}}$$

τ is determined by Type 1 error bound.

$$P[\text{error}] = \overbrace{0.5\int_{-\infty}^{\tau} f_1(x)\,dx}^{e_0} + \overbrace{0.5\int_{\tau}^{\infty} f_0(x)\,dx}^{e_1}$$

$$= 0.5\left[\phi\left(\frac{\tau-\mu_1}{\sigma}\right) + \left\{1-\phi\left(\frac{\tau-\mu_0}{\sigma}\right)\right\}\right]$$

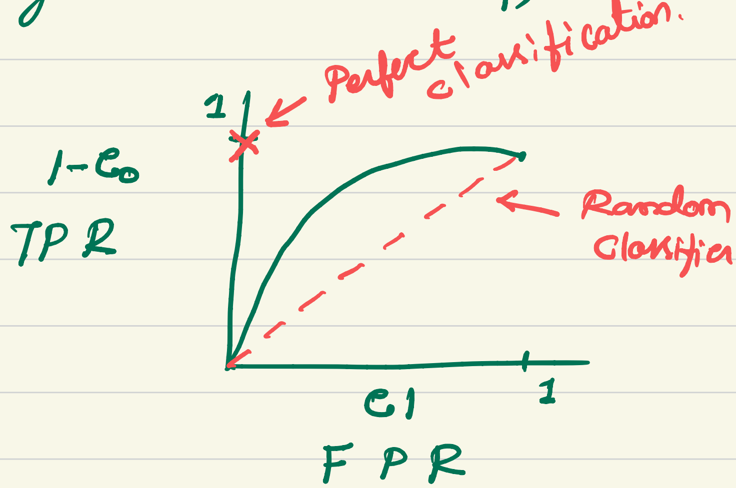
	+ve	-ve ← Actual
+ve	TP : $1-e_0$ ✓ $P[x>\tau \mid x\in C_1]$	False +ve : e_1 $P[x>\tau \mid x\in C_0]$ ✓
-ve	False -ve : e_0 $P[x\le\tau \mid x\in C_1]$ ✓	TN : $1-e_1$ $P[x\le\tau \mid \cancel{x}\in C_0]$ ✓

↑ Predicted

Precision : $\dfrac{TP}{TP+FP}$

Recall = $\dfrac{TP}{TP+FN}$

Vary τ : e_1, $1-e_0$



fix all class cond^n
vary τ point
$(e_1, 1-e_0)$ moves on
Smooth Curve.

" Reciever operating characteristic "

$1-e_0 > e_1 \Rightarrow e_0 + e_1 < 1$

AUC : Area under the curve.

$$\left(\frac{\tau - \mu_0}{\sigma}\right) = \phi^{-1}(1 - e_1) \quad : A$$

τ

$$\left(\frac{\tau - \mu_1}{\sigma}\right) = \phi^{-1}(1-(1-e_0)) \quad : B$$

$$|A - B| = \frac{|\mu_0 - \mu_1|}{\sigma} = C \quad \Big\} \text{ independent of } \tau$$

* if you know C you can get e_1 & $1-e_0$

* tweak τ to get Bayes error.

Tutorial 3

Que- 2-class classif^n

x - feature vector

$y \in \{-1, +1\}$ - class label

$$J(W) = \mathbb{E}_{XY}\left[(W^T X - Y)^2\right]$$

min. is W^*

$\cancel{(X, Y)}$ $(X, \tilde{Y})$ Given

We don't have these

may/may not be same as Y

$(x, \tilde{y})$:-

$$\tilde{y} = \begin{cases} y, & \text{w.p. } 1-\eta \\ \cancel{|1-y|} \; -y, & \text{w.p. } \eta \end{cases}$$

$y \in \{+1, -1\}$

"Randomly flipping labels with prob. η"

Since data is noisily labelled, we can minimize

$$J'(W) = \mathbb{E}_{X\tilde{Y}}\left[(W^T X - \tilde{y})^2\right]$$

$\downarrow$

min. $\tilde{W}^*$

Que- What's the rel^n b/w P_{XY} and $P_{X\tilde{Y}}$?

$\longrightarrow$ ~~No relation b/c labels are randomly flipped.~~

$$f_{XY}(x,y) \qquad\qquad f_{X\tilde{Y}}(x,\tilde{y})$$

$$\mathbb{P}[X=x, Y=y] \qquad\qquad \mathbb{P}[X=x, \tilde{Y}=\tilde{y}]$$

$$f_{X\tilde{Y}}(x,\tilde{y}) \overset{(M)}{=} \sum_y f_{XY\tilde{Y}}(x,y,\tilde{y})$$

$$\overset{(CP)}{=} \sum_y f_{XY}(x,y) f_{\tilde{Y}|X,Y}(\tilde{y}|x,y)$$

$$= \underbrace{f_{XY}(x,y)(1-\eta)}_{y=\tilde{y}} + \underbrace{f_{XY}(x,\tilde{y})(\eta)}_{y\neq\tilde{y}}$$

$$= f_{XY}(x,y)(1-\eta) + \eta f_{XY}(x,\tilde{y})$$

$$\mathbb{E}_{XY}[\ldots] = \mathbb{E}_{X\tilde{Y}}[\ldots]$$

$\boxed{\text{Q} \cdot \text{ue}}$ For any random sample (x, y) drawn acc. to distⁿ P_{XY}, show that the prob. of misclassificⁿ of W^* and $\tilde{W}^*$ are the same.

$$\rightarrow \mathbb{P}_{XY}[\operatorname{sgn}(W^{*T}X) = Y] \overset{?}{=} \mathbb{P}_{XY}[\operatorname{sgn}((\tilde{W}^*)^T X) = Y]$$

$$\min_W \left\{ J(W) = \mathbb{E}_{XY}[(X^T W - Y)^2] \right\}$$

FONC :- $\nabla_W J(W) = 0$

$$\Rightarrow \mathbb{E}_{XY}[X(X^T W - Y)] = 0$$

$$\Rightarrow \quad w^* = \left(\mathbb{E}_x[xx^T]\right)^{-1} \mathbb{E}_{xy}[xy]$$

Similarly,

$$\min_{\tilde{w}} \underline{J(\tilde{w})}$$

$$\downarrow$$

$$\mathbb{E}_{x\tilde{y}}[(x^T\tilde{w} - \tilde{y})^2]$$

$$= \mathbb{E}_x \mathbb{E}_{\tilde{y}|x}[(x^T\tilde{w} - \tilde{y})^2]$$

$$= \mathbb{E}_x \mathbb{E}_{y|x} \mathbb{E}_{\tilde{y}|y,x}[(x^T\tilde{w} - \tilde{y})^2]$$

$$\boxed{\begin{aligned} f(\tilde{y}|x) &= \sum_y f(\tilde{y}, y|x) \\ &= \sum_y f(\tilde{y}|y,x) f(y|x) \end{aligned}}$$

$$= \mathbb{E}_{x} \mathbb{E}_{\eta|x} \Big[(1-\eta)(x^T\tilde{w} - y)^2 + \eta (x^T \tilde{w} + y)^2 \Big]$$

FONC :- $\nabla_{\tilde{w}} J(\tilde{w}) = 0$

$$\Rightarrow \mathbb{E}_{xy} \Big[(1-\eta) x (x^T \tilde{w} - y) + \eta\, x (x^T \tilde{w} + y) \Big] = 0$$

$$\Rightarrow \mathbb{E}_{xy} \big[x(x^T \tilde{w}) - (1-2\eta) xy \big] = 0$$

$$\Rightarrow \tilde{w}^* = (1-2\eta) \left(\mathbb{E}_{x} [x x^T] \right)^{-1} \times \mathbb{E}_{xy} [xy]$$

$$\Rightarrow \boxed{\tilde{w}^* = (1-2\eta) w^*}$$

($\eta < 0.5 \leftarrow$ assume)

$\Downarrow$

$1-2\eta > 0$

For any (x, y) pair, (y: clean)

$$\boxed{\begin{aligned} &\mathrm{sgn}(\tilde{w}^{*T} x) \\ &= \qquad\qquad (\because (1-2\eta) > 0) \\ &\mathrm{sgn}(w^{*T} x) \end{aligned}}$$

$\Downarrow$

Prob of miscla.. under w^* & $\tilde{w}^*$ is the SAME!

(DC signal in noise)

Que: $y = a + w$ $\qquad w \sim \mathcal{N}(0, \sigma^2)$

Assume uniform (0-1) loss

i.e. $L_{00} = L_{11} = 0$

$L_{01} = L_{10} = 1$

$$\ln(L(y)) = \frac{a}{2\sigma^2}(2y-a)$$

LRT:-

$$\frac{a}{2\sigma^2}(2y-a) \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\gamma)$$

$$\Longleftrightarrow \quad y \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{a/2 + \sigma^2/a \ln\gamma}_{\eta}$$

(True Detectⁿ prob.)

$$P_D = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\eta}^{\infty} e^{-(y-a)^2/2\sigma^2}\, dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\frac{\eta-a}{\sigma}}^{\infty} e^{-z^2/2}\, dz$$

$$= Q\left(\frac{\eta - a}{\sigma} \right)$$

$$= Q\left(\frac{\ln \gamma}{d} - \frac{d}{2} \right)$$

$d = a/\sigma$ "SNR" (Signal-to-Noise Ratio)

~

$$P_F = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\eta}^{\infty} e^{-y^2/2\sigma^2}$$

(= False +ve prob.)

$$= Q(\eta/\sigma)$$

$$= Q\left(\frac{\ln \gamma}{d} + \frac{d}{2}\right)$$

Que:- Design minmax classifier for detecting DC signal in noise

$$\boxed{y = a + w \quad w \sim \mathcal{N}(0, \sigma^2)}$$

uniform costs:- $L_{00} = 0 = L_{11}$

$L_{10} = 1 = L_{01}$

$\downarrow$ (minimax rule)

Select prior $\hat{p}_1$ s.t.

$$P_F(\hat{p}_1) = 1 - P_D(\hat{p}_1)$$

$$\Updownarrow$$

$$Q\left(\frac{\ln \gamma}{d} + \frac{d}{2}\right) = 1 - Q\left(\frac{\ln \gamma}{d} - \frac{d}{2}\right)$$

$$\Rightarrow \frac{\ln \gamma}{d} + \frac{d}{2} = -\frac{\ln \gamma}{d} + \frac{d}{2}$$

$$\left(\because 1 - Q(x) = Q(-x)\right)$$

$$\Longleftrightarrow \frac{\ln \gamma}{d} = 0$$

$$\Longleftrightarrow \gamma = 1 \qquad \begin{pmatrix} d \neq 0 \\ \text{as } a \neq 0 \\ \sigma \neq 0 \end{pmatrix}$$

$$\hat{p}_1 = \frac{L_{10} - L_{00}}{(L_{10} - L_{00}) + \gamma(L_{01} - L_{11})}$$

$$= \frac{1 - 0}{1 - 0 + \gamma(1 - 0)} = \frac{1}{2}$$

So, take $\hat{p}_1 = 0.5$ as the worst case prior and continue to use the Bayes classifier (LRT rule) with $\hat{p}_1 = 0.5$

This'll be the desired MINMAX classifier.