۱۳۹۶/ /

سوال ۱

a) غلط. زمانی که تعداد داده ها کم باشد، مدل های پیچیده به راحتی دچار overfitting می شوند.

b) درست، مدل های ساده تر به ازای دیتاست های مختلف کمتر تغییر می کنند ولذا واریانس کمتری دارند ولی خطای بایاس بیشتری دارند.

c) درست، زیرا در تخمین گرهای بیزی از دانش پیشین می توان استفاده کرد و فقط وابسته به داده ها نیست.

d) درست:

$$I(X;Y)=0 \Leftrightarrow X \perp Y \Rightarrow E[XY]=E[X]\,E[Y]$$

e) درست، زیرا منظم ساز باعث کاهش پیچیدگی مدل می شود.

f) درست، زیرا توزیع احتمال هر دو کلاس گاوسی با ماتریس کوواریانس I می شود. چون ماتریس های کوواریانس یکسان است، لذا مرز جداکننده خطی می شود.

g) غلط - الگوریتم پرسپترون همواره مرز خطی جداکننده را در صورت وجود می یابد.

سوال ۲

۱۳۹۶/ /

الف)

$$\log P(D|\lambda) = \log \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \sum_{i=1}^{n} \left(x_i \log \lambda - \lambda - \log x_i! \right)$$

$$= \log \lambda \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log x_i!$$

$$\frac{\partial \log P(D|\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \Rightarrow \boxed{\hat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

ب)

$$P(\lambda|D) = \frac{P(D|\lambda) P(\lambda)}{P(D)} \propto P(D|\lambda) P(\lambda)$$

$$= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \, c \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= c' \lambda^{\sum x_i + \alpha - 1} e^{-(n+\beta)\lambda}$$

$$= c' \lambda^{\alpha'-1} e^{\beta'\lambda} = \mathrm{Gamma}(\lambda | \alpha', \beta')$$

$$\alpha' = \sum_{i=1}^{n} x_i + \alpha \qquad \beta' = n + \beta$$

ج) بله، زیرا توزیع پسین از جنس توزیع پیشین شده است.

د)

$$\lambda_{MAP} = \arg\max_{\lambda} P(\lambda|D) = \frac{\alpha'-1}{\beta'}$$

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} x_i + \alpha - 1}{n+\beta} = \frac{\sum x_i}{n} = \hat{\lambda}_{ML}$$

هـ) بله، زیرا:

سوال ۳

الف) $L(w) = \sum_{i=1}^{n} (e^{wx_i} - y_i)^2$

ب)

$$\frac{\partial L}{\partial w} = \sum_{i=1}^{n} 2x_i e^{wx_i} (e^{wx_i} - y_i)$$

$$\Rightarrow w_{t+1} = w_t - 2\eta \sum_{i=1}^{n} x_i e^{w_t x_i} (e^{w_t x_i} - y_i)$$

ج) d

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \sum_{i=1}^{n} x_i e^{2wx_i} = \sum_{i=1}^{n} y_i x_i e^{wx_i}$$

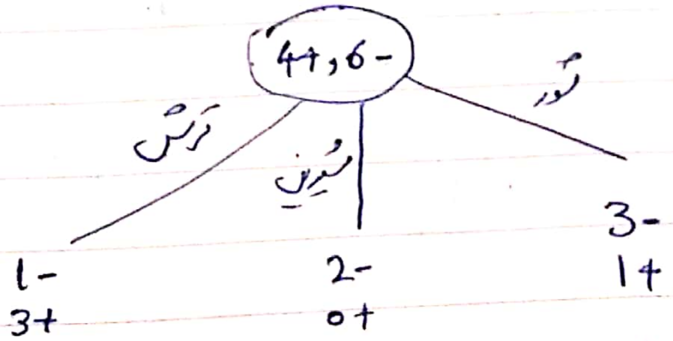
سوال ۴

الف)

$$P(y = \text{خوب}) = \frac{4}{10}$$

$$\Rightarrow H(y) = -\frac{4}{10} \log \frac{4}{10} - \frac{6}{10} \log \frac{6}{10}$$

ب)



$P(y=+ \mid \text{ترش}) = \frac{3}{4}$ $\quad$ $P(y=+ \mid \text{شیرین}) = 0$ $\quad$ $P(y=+ \mid \text{شور}) = \frac{1}{4}$

۱۳۹۶/ /

$$-\frac{3}{4}\log\frac{3}{4}-\frac{1}{4}\log\frac{1}{4}$$

$$H(y|\text{طعم})=\frac{4}{10}H\left(\frac{3}{4}\right)+\frac{2}{10}\overbrace{H(0)}^{0}+\frac{4}{10}H\left(\frac{1}{4}\right)$$

$$I(y;\text{طعم})=H(y)-H(y|\text{طعم})$$

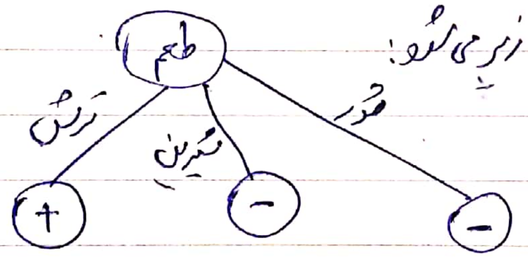
$$=-\frac{4}{10}\log\frac{4}{10}-\frac{6}{10}\log\frac{6}{10}-\frac{8}{10}\left(-\frac{3}{4}\log\frac{3}{4}-\frac{1}{4}\log\frac{1}{4}\right)$$

راه دوم: مستقیما از رابطه ی زیر هم می توان استفاده کرد:

$$H(y|x)=\sum_{x,y}P(x,y)\log P(y|x)$$

ج)

درخت تصمیم به صورت زیر می شود:



کلاس واقعی \ کلاس پیش بینی	+	−
+	TP=3	FN=1
−	FP=1	TN=5

$P=4$

$N=6$

$$\text{recall}=\frac{TP}{P}=\frac{3}{4}\qquad \text{precision}=\frac{TP}{TP+FP}=\frac{3}{4}$$

$$\text{specificity}=\frac{TN}{N}=\frac{5}{6}\qquad \text{acc}=\frac{TP+TN}{n}=\frac{8}{10}$$

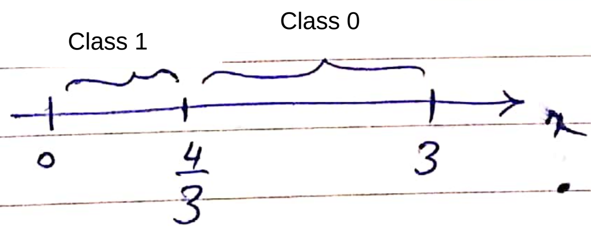
سوال ۵   مرز از رابطه‌ی زیر بدست می‌آید :   ۱۳۹۶/ /

$$\frac{P(y=0|x)}{P(y=1|x)} = \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}.$$

$$\cancel{\Rightarrow \frac{P(x|y=0)}{P(x|y=1)}}$$

$$\Rightarrow \frac{\dfrac{P(x|y=0)\,P(y=0)}{P(x)}}{\dfrac{P(x|y=1)\,P(y=1)}{P(x)}} = \frac{P(x|y=0)}{P(x|y=1)} = \frac{x^2/9}{2x/9} = \frac{x}{2}$$

$$\Rightarrow \frac{x}{2} = \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} = \frac{4}{6}. \Rightarrow \boxed{x = \frac{4}{3}}$$



سوال ۶

الف - overfitting ← راه‌حل : drop out

Stochastic Gradient Descent

ب)

$n \to \infty$

$k \to \infty$

$k/n \to 0$