ML Exam Retake 20

1A (a&b) A linear classifier makes classification decision based on the value of a linear combination of the characteristics. There are two broad classes; generative and discriminative. Examples of generative models are LDA and Naive Bayes Classifier. Examples of descriminative models are Logistic Regression, Perceptron, and Support Vector Machines.

(~~b~~ c) Generative models classify ~~based on the~~ after learning joint probability distribution ~~conditional~~ density ~~functions~~ e.g. Gaussian (LDA), Multinomial / Multivariate (NBC). Discriminative models directly predict the conditional probabilities based on the training data.

1B (A) $\bar{x}_A = -0.5$, $\bar{Y}_A = 0$, $\bar{x}_B = 1$, $\bar{Y}_B = 0.33$, $\hat{n}_B = 3$

$\pi_B = 3/7$

$\hat{n}_A = 4$, $\pi_A = 4/7$

$$\cancel{\Sigma_A = \frac{1}{4}\begin{bmatrix} -0.5 & 1 \\ -0.5 & -1 \\ 0.5 & 1 \\ 0.5 & -1 \end{bmatrix}^T \begin{bmatrix} -0.5 & -0.5 & 0.5 & 0.5 \\ 1 & -1 & 1 & -1 \end{bmatrix} = \frac{1}{4}}$$

$$\Sigma_A = \frac{1}{4}\begin{bmatrix} -0.5 & -0.5 & 0.5 & 0.5 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} -0.5 & 1 \\ -0.5 & -1 \\ 0.5 & 1 \\ 0.5 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma_B = \frac{1}{3}\begin{bmatrix} -1 & 1 & 0 \\ -0.33 & -0.33 & 0.67 \end{bmatrix} \begin{bmatrix} -1 & -0.33 \\ 1 & -0.33 \\ 0 & 0.67 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & 0 \\ 0 & 0.67 \end{bmatrix} = \begin{bmatrix} 0.67 & 0 \\ 0 & 0.22 \end{bmatrix}$$

(B)

$$\Sigma = \frac{4}{7}\Sigma_A + \frac{3}{7}\Sigma_B = \begin{bmatrix} 0.43 & 0 \\ 0 & 0.67 \end{bmatrix}$$

$$\delta_A(x) = -\frac{1}{2}\log|\Sigma_A| - \frac{1}{2}\langle x-\bar{x}_A, \Sigma^{-1}(x-\bar{x}_A)\rangle + \log \pi_A,$$

$$\delta_A(x) = -\frac{1}{2}\log(0.29) - \frac{1}{2}\left\langle \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}, \begin{bmatrix} \overset{1.16}{\cancel{2}} \\ 0 \end{bmatrix} \right\rangle + \log\left(\frac{4}{7}\right) = +0.\overset{62}{\cancel{\ }} - 0.\overset{29}{\cancel{\ }} - 0.56 = \overset{-0.23}{\cancel{-1.25}}$$

$$\delta_B(x) = -\frac{1}{2}\log(0.29) - \frac{1}{2}\left\langle \begin{bmatrix} -1 \\ -0.33 \end{bmatrix}, \begin{bmatrix} -2.33 \\ -0.49 \end{bmatrix} \right\rangle + \log\left(\frac{3}{7}\right) = \overset{0.62}{\cancel{0.06}} - 1.246 - 0.85 = \overset{-1.476}{\cancel{-2.036}}.$$

$$\operatorname{argmax}(\delta_A(x), \delta_B(x)) = \operatorname{argmax}(\overset{-0.23,\ -1.476}{\cancel{-1.25, -2.036}}) = \text{Class A}$$

1(c) (i) In LDA, decision boundaries are linear while in QDA, these are quadratic.

(ii) 2a) QDA 2b) QDA 2c) Both are inappropriate.

2A. 1) Numerical Variables — Euclidean Distance

2) Nominal variables — ~~&~~ L_∞, $\frac{1}{2}L_1$, $\frac{1}{\sqrt{2}}L_2$ — Slide 10

3) Set-valued variables — Hamming Distance, Jaccard Coefficient.

4) String variables — Levenshtein Distance

2B, (i) entre'e, entreat, entity, entomb.

		s	w	a	p
	0	1	2	3	4
s	1	0	1	2	3
w	2	1	0	1	2
p	3	2	1	1	1
a	4	3	2	1	$\boxed{2}$

Since many words will have the same distance from the word 'lie', we can add a feature to the dictionary, based on which we can add a weight to ~~the~~ ~~dist~~ each word and its distance. In this way, we will use weighted distances. Using such a model, words having meaning similar to 'lie' will be selected. ~~which could be~~ Another way could be to add a 'part of speech' feature (verb, noun, adjective etc.). to the model, which will help in selecting more suitable words.

2C) We can use Lower Bounding, Locality sensitive Hashing or Editing techniques.

3A) (a) $p(a,b,c,w,x,y,z) = p(a).p(b).p(c).p(w|a,b,c).p(x|w).p(y|w).p(z|w)$

(b) $p(a,b,c,w,x,y,z) = p(a).p(w).p(b|a).p(c|a,w).p(x|b).p(y|c).p(z|x,y)$.

(c) $p(a,\cancel{b,c},w,x,y,z) = p(a).p(x|a).p(y|a,x).p(z|a,y).p(w|x,y,z)$

ML Exam Retake 2020

N95.

TP	N_1	N_0	θ_1	θ_0
			1/4	3/4
No	0	2		
			$\cancel{3}4/6$	2/6
Yes	3	1		

Pasta

TP	N_1	N_0	θ_1	θ_0
			2/4	2/4
No	1	1		
			$\cancel{3}4/6$	2/6
Yes	3	1		

Sanitizer

TP	N_1	N_0	θ_1	θ_0
			2/4	2/4
No	1	1		
			$\cancel{3}4/6$	2/6
Yes	3	1		

So, $P(Y|Y,Y,Y) = 0.94 \rightarrow$ Predicted Class

$P(N|Y,Y,Y) = 0.06$ Yes.

$$P(Y|Y,Y,Y) = \frac{P(Y|N95_Y)\cdot P(Y|P_Y)\cdot P(Y|S_Y)\ P(Y)}{\sum_z P(Y,Y,Y,Y)}$$

$$= \frac{(\frac{4}{6})(\frac{4}{6})(\frac{4}{6})(\frac{4}{6})}{(\frac{4}{6})(\frac{4}{6})(\frac{4}{6})(\frac{4}{6}) + (\frac{2}{6})(\frac{2}{6})(\frac{2}{6})(\frac{2}{6})} = 0.94$$

$$P(N|Y,Y,Y) = \frac{P(N|N95_Y)\cdot P(N|P_Y)\cdot P(N|S_Y)\cdot P(N)}{\sum_z P(N,Y,Y,Y)}$$

$$= \frac{(\frac{2}{6})(\frac{2}{6})(\frac{2}{6})(\frac{2}{6})}{(\frac{4}{6})(\frac{4}{6})(\frac{4}{6})(\frac{4}{6}) + (\frac{2}{6})(\frac{2}{6})(\frac{2}{6})(\frac{2}{6})} = 0.058$$