JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740 Fall '20

Final Examination

From the time you download the exam, you have four (4) hours to complete it and turn it in. The latest time to turn in the exam is December 9 at 12:01 am PST. Please upload your exam to `blackboard.jhu.edu`
.

Please type up your answers. If you are not able to type your answers, please write very neatly. Sloppy, small, or otherwise difficult-to-read writing will not be accepted. Please include in your submission all code in electronic form.

You may use the textbook Pattern Classification, by Duda, et al. You may use the material posted to Blackboard: notes, videos, textbook chapters. No other sources may be consulted.

It should go without saying that posting questions to or viewing questions at "homework and exam" solving websites and the like is not allowed. Searching the internet, or obtaining notes from other classes is not allowed on this exam. During the time of the exam, you may discuss these problems only with the course instructor and nobody else.

Time and date of exam start: 4:48 12/7/20

Time and date of exam finish: ____________________

This exam represents my own work. I did not consult with anybody during the time of the exam. I did not look up any of these problems on or post any of these problems to the internet.

Signature: ____________________

Name: Jared Yu

Date: 12/7/20

1. Suppose we have two normal distributions with the same covariances but different means: $\mathcal{N}(\boldsymbol{\mu_1}, \boldsymbol{\Sigma})$ and $\mathcal{N}(\boldsymbol{\mu_2}, \boldsymbol{\Sigma})$. In terms of the prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary will pass between the two means.

2. Describe in detail when a Bayesian Classifier will have a decision boundary given by
   (a) a circle,
   (b) an ellipse,
   (c) an hyperbola,
   (d) a parabola,
   (e) a straight line.

3. (a) Given four points not all coplanar, how many planes can be drawn, equidistant from all the points (and thus separate these points into two classes)?
   (b) Given five points not all cohyperplanar, how many hyperplanes can be drawn, equidistant from all the points (and thus separate these points into two classes)?

4. Let p_n=parity$(x_1, x_2, \ldots, x_n)$.
   (a) Show a neural network that calculates p_{16}.
   (b) Show a neural network that calculates p_{127}.
   (c) Write down the gradient descent update equations for a network to calculate p_8.
   (d) Train such a neural network and show weights and the output for each of
      i. 10101010,
      ii. 11000110,
      iii. 10001000,
      iv. 11111111,

5. Consider three Gaussian pdfs: $X_1 \sim \mathcal{N}(1, 0.1), X_2 \sim \mathcal{N}(2, 0.1), X_3 \sim \mathcal{N}(3, 0.2)$, with *prior* probabilities given by $P_1 = 1/6, P_2 = 1/2, P_3 = 1/3$. The pdf underlying the random samples is modeled as a mixture
$$\sum_{j=1}^{3} \mathcal{N}(\mu_j, \sigma_j^2) P_j.$$
   Generate 500 samples according to this model and now, assume the parameters of the mixture are unknown. Use the EM algorithm and these samples to estimate the unknown parameters μ_j, σ_j^2, P_j.

(1) Let w_1 denote class for $N(\mu_1, \Sigma)$

" w_2 " " " $N(\mu_2, \Sigma)$

$\vec{x}$ is d-dimension obs.

$p(\vec{x}, w)$ is class-cond. probability

$$\vec{x} \mid w = w_1 \sim N_d(\mu_1, \Sigma)$$

$$\vec{x} \mid w = w_2 \sim N_d(\mu_2, \Sigma)$$

$$p(\vec{x} \mid w_1) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\mu_1)^T \Sigma^{-1} (\vec{x}-\mu_1)\right]$$

$$p(\vec{x} \mid w_2) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\mu_2)^T \Sigma^{-1} (\vec{x}-\mu_2)\right]$$

classify an obs. to w_1 if

$$P(w_1 \mid \vec{x}) > P(w_2 \mid \vec{x})$$

$$\frac{P(w_1 \mid \vec{x})}{P(w_2 \mid \vec{x})} > 1 \Rightarrow \frac{p(\vec{x} \mid w_1) P(w_1) / p(\vec{x})}{p(\vec{x} \mid w_2) P(w_2) / p(\vec{x})} > 1$$

$$\frac{\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)\right]}{\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\mu_2)^T\Sigma^{-1}(\vec{x}-\mu_2)\right]} > \frac{P(\omega_2)}{P(\omega_1)}$$

$$\exp\left[-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)\right] > \exp\left[-\frac{1}{2}(\vec{x}-\mu_2)^T\Sigma^{-1}(\vec{x}-\mu_2)\right]\frac{P(\omega_2)}{P(\omega_1)}$$

$$-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1) > -\frac{1}{2}(\vec{x}-\mu_2)^T\Sigma^{-1}(\vec{x}-\mu_2) + \ln P(\omega_2) - \ln P(\omega_1)$$

$$(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1) < (\vec{x}-\mu_2)^T\Sigma^{-1}(\vec{x}-\mu_2) - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$(\vec{x}^T\Sigma^{-1} - \mu_1\Sigma^{-1})(\vec{x}-\mu_1) < (\vec{x}^T\Sigma^{-1} - \mu_2^T\Sigma^{-1})(\vec{x}-\mu_2) - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$\vec{x}^T\Sigma^{-1}\vec{x} - \mu_1^T\Sigma^{-1}\vec{x} - \vec{x}^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}\mu_1 <$$

$$\vec{x}^T\Sigma^{-1}\vec{x} - \mu_2^T\Sigma^{-1}\vec{x} - \vec{x}^T\Sigma^{-1}\mu_2 + \mu_2^T\Sigma^{-1}\mu_2 - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

for $i = 1, 2$

$$\vec{x}^T\Sigma^{-1}\mu_i = (\vec{x}^T\Sigma^{-1}\mu_i) = \mu_i^T\Sigma^{-1}\vec{x}$$

$$\Rightarrow -2\vec{x}^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}\mu_1 < -\vec{x}^T\Sigma^{-1}\mu_2 + \mu_2^T\Sigma^{-1}\mu_2 - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

Then classify w_1 if:

$$2\vec{x}^T\Sigma^{-1}(\mu_2-\mu_1) < \mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1 - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

Then classify w_2 if:

$$\vec{x}^T\Sigma^{-1}(\mu_2-\mu_1) < \frac{\mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1}{2} - 2\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

(2) Generally, assume

$$\vec{x}|w_1 \sim N(\mu_1, \Sigma_1) \quad \text{w/ prior } p(w_1)$$

$$\vec{x}|w_2 \sim N(\mu_2, \Sigma_2) \quad " \quad p(w_2)$$

classify to w_1 if

$$p(w_1|\vec{x}) > p(w_2|\vec{x})$$

$$\frac{p(\vec{x}|w_1)}{p(\vec{x}|w_2)} > \frac{p(w_2)}{p(w_1)}$$

$$\frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{\exp\left[-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma_1^{-1}(\vec{x}-\mu_1)\right]}{\exp\left[-\frac{1}{2}(\vec{x}-\mu_2)^T\Sigma_2^{-1}(\vec{x}-\mu_2)\right]}$$

$$-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma_1^{-1}(\vec{x}-\mu_1) - \frac{1}{2}\ln|\Sigma_1| + \ln p(w_1)$$
$$> -\frac{1}{2}(\vec{x}-\mu_2)^T\Sigma_2^{-1}(\vec{x}-\mu_2) - \frac{1}{2}\ln|\Sigma_2| + \ln p(w_2)$$

$$-\frac{1}{2}\left[\vec{x}^T\Sigma_1^{-1}\vec{x} - 2\mu_1^T\Sigma_1^{-1}\vec{x} + \mu_1^T\Sigma_1^{-1}\mu_1\right] - \frac{1}{2}\ln|\Sigma_1| + \ln p(w_1)$$
$$> -\frac{1}{2}\left[\vec{x}^T\Sigma_2^{-1}\vec{x} - 2\mu_2^T\Sigma_2^{-1}\vec{x} + \mu_2^T\Sigma_2^{-1}\mu_2\right] - \frac{1}{2}\ln|\Sigma_2| + \ln p(w_2)$$

$$\vec{x}^T\left[-\frac{1}{2}(\Sigma_1^{-1} - \Sigma_2^{-1})\right]\vec{x} + (\mu_1^T\Sigma_1^{-1} - \mu_2^T\Sigma_2^{-1})\vec{x} +$$
$$\left[-\frac{1}{2}\mu_1^T\Sigma_1^{-1}\mu_1 - \frac{1}{2}\ln|\Sigma_1| + \ln p(w_1)\right] -$$
$$\left[-\frac{1}{2}\mu_2^T\Sigma_2^{-1}\mu_2 - \frac{1}{2}\ln|\Sigma_2| + \ln p(w_2)\right] > 0$$

Let $w_{i0} = -\frac{1}{2}\mu_i^T\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln p(w_i)$

Then we classify $\vec{x}$ to class w_1 if:

$\vec{x}^T[-\frac{1}{2}(\Sigma_1^{-1}-\Sigma_2^{-1})]\vec{x}+(\mu_1^T\Sigma_1^{-1}-\mu_2^T\Sigma_2^{-1})\vec{x}$ +

$(w_{10}-w_{20}) \geq 0$

The Bayesian classifier will have the following decision boundary

$$g(\vec{x}) = \vec{x}^T[-\frac{1}{2}(\Sigma_1^{-1}-\Sigma_2^{-1})]\vec{x}+(\mu_1^T\Sigma_1^{-1}-\mu_2^T\Sigma_2^{-1})\vec{x}+ (w_{10}-w_{20})$$

In order to explain it in details, we consider two features cases:

i.e., $\vec{x}|w_1 \sim N(\mu_1, \Sigma_1)$ w/ prior $p(w_1)$

$\vec{x}|w_2 \sim N(\mu_2, \Sigma_2)$ " $p(w_2)$

2a) We have a circle decision boundary if $-\frac{1}{2}(\Sigma_1^{-1}-\Sigma_2^{-1})$ in $\vec{x}^T[-\frac{1}{2}(\Sigma_1^{-1}-\Sigma_2^{-1})]\vec{x}$ is in aI form for some constant a. In other words if $\Sigma_1^{-1}=\sigma_1^2 I$, $\Sigma_2^{-1}=\sigma_2^2 I$, $\sigma_1 \neq \sigma_2$

$$-\frac{1}{2}(\Sigma_1^{-1}-\Sigma_2^{-1}) = -\frac{1}{2}(\sigma_1^2-\sigma_2^2)I$$

Then, the boundary is a circle

2b) We have an ellipse decision boundary if

$-\frac{1}{2}(\Sigma_1^{-1} - \Sigma_2^{-1})$ is a diagonal matrix but not all

the diagonal elements are the same.

For example, if $p(w_1) = p(w_2)$, $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\mu_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ $\Sigma_1 = \begin{bmatrix} 0.31 & 0 \\ 0 & 0.36 \end{bmatrix}$ $\Sigma_2 = \begin{bmatrix} 1.2 & 0 \\ 0 & 1.8 \end{bmatrix}$

Then we get ellipse decision boundary.

2c) Combining from (a) and (b), we

can see that $-\frac{1}{2}(\Sigma_1^{-1} - \Sigma_2^{-1})$ is the key to

determine the shape of the boundary.

If we want an hyperbola decision boundary,

we need two equiprobable sets of orthogonal

elliptical isocontours.

For example, $p(w_1) = p(w_2)$, $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$,

$\Sigma_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.7 \end{bmatrix}$ $\Sigma_2 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.1 \end{bmatrix}$ in this case

the decision boundary is hyperbolic.

2d) We will have parabolic decision boundary if one distribution has a circular isocontour.

For example, $\Sigma_1 = \sigma_1^2 I$ and Σ_2 is diagonal matrix w/ different values in diagonal.

2e) If $\Sigma_1 = \Sigma_2 = \Sigma$

$$g(\vec{x}) = \vec{x}^T \left[-\frac{1}{2} \left(\Sigma_1^{-1} - \Sigma_2^{-1} \right) \right] \vec{x} + \left(\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1} \right) \vec{x} + (w_{10} - w_{20})$$

$$= \vec{x}^T \left[-\frac{1}{2} \left(\Sigma^{-1} - \Sigma^{-1} \right) \right] \vec{x} + \left(\mu_1^T \Sigma^{-1} - \mu_2^T \Sigma^{-1} \right) \vec{x} + (w_{10} - w_{20})$$

$$= (\mu_1^T - \mu_2^T) \Sigma^{-1} \vec{x} + w_{10} - w_{20}$$

we get linear decision boundary.

3a) Let the four points are A, B, C, D

Class 1	Class 2
A	BCD
B	ACD
C	ABD
D	ABC
AB	CD
AC	BD
AD	BC

So, total 7 equidistant planes can be drawn to separate these points into two classes.

4 ways to separate one point from the rest

3 " " " two points " " "

So, total 7 equidistant planes to do it.

3b) Let A, B, C, D, E be the 5 points

class 1	class 2	
A	BCDE	5 ways
B	ACDE	
C	ABDE	
D	ABCE	
E	ABCD	
AB	CDE	4 ways
AC	BDE	
AD	BCE	
AE	BCD	
BC	ADE	3 ways
BD	ACE	
BE	ACD	
CD	ABE	2 ways
CE	ABD	
DE	ABC	1 ways

There are 15 equidistant hyperplanes can be drawn to separate these points into two classes.