# MIDTERM EXAM

Math 251: Statistical and Machine Learning Classification
Instructor: Guangliang Chen
San Jose State University
Fall 2018

*You have 75 minutes.*

*No books or notes are allowed.*

*Please write legibly (unrecognizable work will receive zero credit).*

*You must show your work to support your answer in order to receive full credit.*

*Good luck!*

Name: ______________________________

1. ____________
2. ____________
3. ____________
4. ____________
5. ____________

"I have adhered to the SJSU Academic Integrity Policy in completing this exam."

Signature: ______________________

Date: ______________________

Total score: ________________ (/60 points)

1. *(10 pts)* Answer the following questions.

(a) What is an orthogonal matrix?

**Answer**. An orthogonal matrix is a square matrix whose inverse is the same as its transpose.

Alternatively, you can say that an orthogonal matrix is a square matrix whose columns are orthonormal vectors.

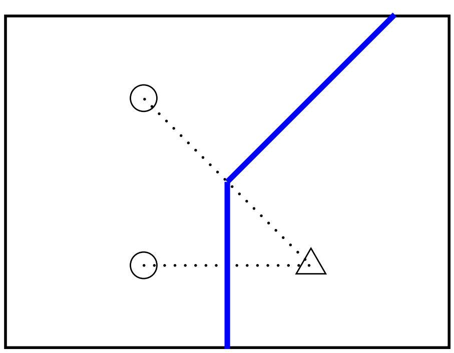
(b) What is a classifier?

**Answer**. A classifier is a rule learned on labeled training data for predicting the labels of previously unseen data.

(c) What is the difference between validation and test errors?

**Answer**. The main difference is that the validation error is evaluated on the training data by using cross validation, while the test error corresponds to test data.

(d) Sketch as accurately as possible the decision boundary of 1NN based on the following training data (2 classes):



2. *(16 pts)* Let $\mathbf{A} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}$

(a) Find the full SVD of $\mathbf{A}$.

**Answer**. We first compute

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & 0 \\ \sqrt{2} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{5}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

Its characteristic polynomial is $\det(\mathbf{A}^T\mathbf{A} - \lambda\mathbf{I}) = (\lambda - \frac{5}{2})^2 - (\frac{3}{2})^2$, from which we obtain two eigenvalues $\lambda_1 = \frac{5}{2} + \frac{3}{2} = 4$ and $\lambda_1 = \frac{5}{2} - \frac{3}{2} = 1$. This implies that there are two distinct singular values $\sigma_1 = 2$ and $\sigma_2 = 1$.

By direct calculation, the corresponding eigenvectors are $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1)^T$ and $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,-1)^T$. These are also the right singular vectors of $\mathbf{A}$.

The left singular vectors are

$$\mathbf{u}_1 = \frac{1}{\sigma_1}\mathbf{A}\mathbf{v}_1 = \frac{1}{2}\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \frac{1}{\sigma_2}\mathbf{A}\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ (by selection).}$$

Putting everything together, we obtain the following full SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & \\ & 1 \\ & \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T.$$

(b) Write down the compact SVD of $\mathbf{A}$.

**Answer**.

$$\mathbf{A} = \begin{pmatrix} 1 & \\ & 1 \\ & \end{pmatrix} \cdot \begin{pmatrix} 2 & \\ & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T.$$

(c) Compute $(\mathbf{A}^T\mathbf{A})^{\frac{1}{2}}$.

**Answer**. In part (a), we have obtained that $(\mathbf{A}^T\mathbf{A})^{\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ with $\mathbf{\Lambda} = \text{diag}(4,1)$ and $\mathbf{V}$ shown above. Thus,

$$(\mathbf{A}^T\mathbf{A})^{\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}^{\mathbf{1/2}}\mathbf{V}^T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & \\ & 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

3. *(12 pts)* Assume a data set of 4 points $(1,1),(2,4),(4,2),(5,5)$ in $\mathbb{R}^2$.

(a) Consider the projection of the data onto the line $y=3$. How much variance is explained by the projections?

**Answer**. The projections in this case are $1,2,4,5$ (with mean $(1+2+4+5)/4=3$). The amount of variance explained by them is $(1-3)^2+(2-3)^2+(4-3)^2+(5-3)^2=10$.

(b) Find the first principal direction as well as the principal components of the data. How much variance does this direction explain?

**Answer**. The mean of the data is $(3,3)$ from which we get the centered data

$$\tilde{\mathbf{A}} = \begin{pmatrix} -2 & -2 \\ -1 & 1 \\ 1 & -1 \\ 2 & 2 \end{pmatrix}.$$

The covariance matrix is $\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$, with characteristic polynomial $(\lambda-10)^2-6^2$. The largest eigenvalue is $\lambda_1=16$, with associated eigenvector $\mathbf{v}_1=\frac{1}{\sqrt{2}}(1,1)^T$ (this is also the first principal direction). The first principal component is

$$\tilde{\mathbf{A}}\mathbf{v}_1 = [-2\sqrt{2},0,0,2\sqrt{2}],$$

and the amount of variance explained is 16.

(c) What is the coordinate of the projection of a new point $(4,3)$ onto the first principal direction?

**Answer**.

$$\mathbf{v}_1^T(\mathbf{x}-\mathbf{m}) = \frac{1}{\sqrt{2}}(1,1)\cdot\left[\begin{pmatrix} 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 3 \end{pmatrix}\right] = \frac{1}{\sqrt{2}}.$$

4. *(12 pts)* Consider a data set consisting of two classes:

- Class 1: (1,2), (2,3), (3, 4.3)
- Class 2: (2,1), (3,2), (4, 3.3)

(a) Find the between-class scatter matrix for the above given data.

**Answer**. The class means are $\mathbf{m}_1 = (2, 3.1)^T$ and $\mathbf{m}_1 = (3, 2.1)^T$, from which we obtain the following between-class scatter matrix

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

(b) What is the distance between the projections of the two class centroids onto the direction $\mathbf{v} = (0.6, 0.8)^T$?

**Answer**. The distance is

$$\sqrt{\mathbf{v}^T \mathbf{S}_b \mathbf{v}} = \sqrt{(0.6, 0.8) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix}} = 0.2.$$

Alternatively, we can first find the projected centrioids $\mu_1 = \mathbf{v}^T \mathbf{m}_1 = 3.68$ and $\mu_2 = \mathbf{v}^T \mathbf{m}_2 = 3.48$, from which we get their distance $|3.68 - 3.48| = 0.2$.

(c) How much variance does the first class have after being projected onto the direction $\mathbf{v} = (1, 0)^T$?

**Answer**. The projections of class 1 onto this direction are $1, 2, 3$, and the amount of variance explained by them is $(1-2)^2 + (2-2)^2 + (3-2)^2 = 2$.

Alternatively, we can start by computing

$$\mathbf{S}_1 = \sum_{\mathbf{x}_i \in C_1} (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T = \begin{pmatrix} 2 & 2.3 \\ 2.3 & 2.66 \end{pmatrix}$$

It follows that the variance of class 1 in the projection space is

$$\mathbf{v}^T \mathbf{S}_1 \mathbf{v} = (1, 0) \begin{pmatrix} 2 & 2.3 \\ 2.3 & 2.66 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2.$$

5. *(10 pts)* Derive an explicit formula of the Bayes decision rule for a mixture model of two 1-dimensional normal distributions $N(-1, 1^2)$ and $N(3, 1^2)$, with sampling frequencies $\pi_1 = \frac{1}{1+e}$ and $\pi_2 = \frac{e}{1+e}$ respectively. What is the boundary point?

**Answer**. The Bayes decision rule is

$$\hat{j} = \arg\max_j \pi_j f_j(x)$$

Since $f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}}$ and $f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}$, we have

$$\hat{j} = \arg\max \left( \frac{1}{1+e} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}},\ \frac{e}{1+e} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} \right)$$

The boundary point is the solution of

$$\frac{1}{1+e} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} = \frac{e}{1+e} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}$$

Simplifying the equation gives that

$$e^{-\frac{(x+1)^2}{2}} = e^{1-\frac{(x-3)^2}{2}}$$

or

$$-\frac{(x+1)^2}{2} = 1 - \frac{(x-3)^2}{2}$$

from which we obtain the boundary point $x = \frac{3}{4}$.