CMSC 25300 / 35300
Practice Exam 1 – Winter 2019

Name: ______________________________

(Each problem is worth 20 points)

1. ____________

2. ____________

3. ____________

4. ____________

5. ____________

____________ **(total score)**

1. While you are working for NASA, they launch n satellites into space. For each satellite, we write its 3d location (relative to Earth) as a vector $\boldsymbol{x}_i \in R^3$, $i = 1, \ldots, n$. For each pair of satellites i and j, the distance between them is

$$d_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2.$$

Form a matrix $D \in \mathbb{R}^{n \times n}$ where the (i, j) element of the matrix is $d_{i,j}^2$. Solve the below problems, *carefully stating any and all assumptions you make.* If we assume all the satellites are roughly equidistant from Earth, so that $\|\boldsymbol{x}_i\|_2 = 1$ for $i = 1, \ldots, n$, what is the maximum rank of D?

SOLUTION: First note

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 = \|\boldsymbol{x}_i\|_2^2 - 2\boldsymbol{x}_i^\top \boldsymbol{x}_j + \|\boldsymbol{x}_j\|_2^2.$$

In this subproblem, $\|\boldsymbol{x}_i\|_2^2 = \|\boldsymbol{x}_j\|_2^2 = 1$. Define

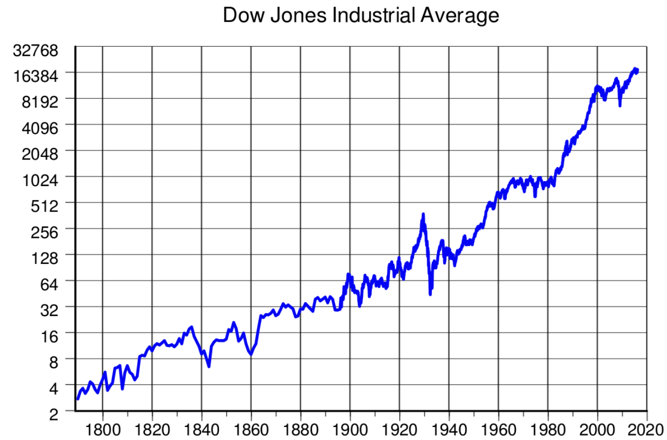
$$X = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} \in \mathbb{R}^{p \times n}.$$

Then we may write

$$D = 2\mathbb{1}_{n \times n} - X^\top X,$$

where $\mathbb{1}_{n \times p}$ is a matrix with n rows and p columns, where every entry has value 1. Now, the rank of X is at most $p = 3$. Let us assume that the p rows of X are linearly independent, which is possible if $n \geq p$ and likely when n is much greater than p, then $\text{rank}(X) = 3$ and hence $rank(X^\top X) = 3$. If $\mathbb{1}_{n \times 1}$ is not in the span of the rows of X, then the rank of D is $p + 1 = 4$.

2. You see the below plot of the Dow Jones Industrial Average stock market index and hypothesize that at time t, the index value is $v_t \approx e^{at+b}$. For simplicity, assume we measure the index value at times $t = 1, 2, \ldots, T$, where $t = 1$ corresponds to 1795 and $t = T$ corresponds to 2017.



a) Explain how you would estimate a and b. Specify any matrices and vectors you would need to use.

SOLUTION: let

$$\boldsymbol{y} = \begin{bmatrix} \log v_1 & \log v_2 & \cdots & \log v_T \end{bmatrix}^\top \quad \text{and} \quad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{bmatrix} \quad \text{and} \quad \boldsymbol{w} = \begin{bmatrix} b \\ a \end{bmatrix}.$$

Our model is $\boldsymbol{y} = X\boldsymbol{w}$. In this case the columns of X are linearly independent, and we can estimate

$$\widehat{\boldsymbol{w}} = (X^\top X)^{-1} X^\top \boldsymbol{y}.$$

b) Partially solve the least squares problem. That is, if you had a least squares problem of the form $(X^\top X)^{-1}X^\top y$, then give expressions for $X^\top X$ and $X^\top y$. HINT: $\sum_{t=1}^T t = \frac{T(T+1)}{2}$ and $\sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$.

SOLUTION: Note

$$X^\top X = \begin{bmatrix} \sum_{t=1}^T 1 & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{bmatrix} = \begin{bmatrix} T & \frac{T(T+1)}{2} \\ \frac{T(T+1)}{2} & \frac{T(T+1)(2T+1)}{6} \end{bmatrix}.$$

Furthermore,

$$X^\top \boldsymbol{y} = \begin{bmatrix} \sum_{t=1}^T \log v_t \\ \sum_{t=1}^T t \log v_t \end{bmatrix}.$$

3. Congratulations! You have been hired by Google's machine learning group. Your first task is to develop a classifier than will indicate whether a web page is political in nature or not. You first write a script that, for any web page, counts the number of times different $p = 6$ keywords (Trump, dog, refugee, clearance, Xbox, and football) appear in the page. You also have an intern who has looked at n different web pages and labeled them as "political" or "apolitical".

a) Explain in detail how you would design your classifier.
SOLUTION: For each of the n web pages, create a length-6 feature vector denoted $\boldsymbol{x}_i$, where the j^{th} element is $\boldsymbol{x}_{i,j}$ is the number of times the j^{th} keyword appeared in the i^{th} web page. Concatenate these into a matrix $X = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}^\top \in \mathbb{R}^{n \times p}$. For a label vector $\boldsymbol{y}$, where the i^{th} element is the label $y_i \in \{-1, +1\}$ indicating whether the i^{th} web page is political. Check to make sure the p columns of X are linearly independent. If so, compute

$$\widehat{\boldsymbol{w}} = (X^\top X)^{-1} X^\top \boldsymbol{y}.$$

For a new web page, we'd compute its word count feature vector $\boldsymbol{x}$ and set

$$\widehat{y} = \langle \boldsymbol{x}, \widehat{\boldsymbol{w}} \rangle$$
$$\tilde{y} = \begin{cases} 1, & \text{if } \widehat{y} \geq 0 \\ -1, & \text{if } \widehat{y} < 0 \end{cases}$$

b) What pitfalls do you need to be wary of?
SOLUTION: Feature vectors should be linearly independent. It may be better to include a constant offset. Need $n \geq p$.

c) How can you predict how well your classifier will work? (Don't just name a technique, but describe it in enough detail to demonstrate you understand it.)
SOLUTION: use k-fold cross-validation as in homework.

d) Write a few lines of matlab or python code illustrating how you would compute your classifier given the training data and how you would classify a new web page.
SOLUTION:

```
hatw = inv(X'*X)*X'*y;
haty = hatw'*xnew;
tildey = 2*(haty>0)-1;
```

4. AllRecipes.com has a large collection of different receipes and ratings their customers have given to each recipe. They hire you to help them analyze this data so they can better help customers find recipes they are likely to enjoy. They start with $p = 5$ recipes – Apple pie, Beef stew, Carrot salad, Devil's food cake, and Eggs benedict. Here are some customer ratings (each column corresponds to a different customer):

$$\begin{bmatrix}4\\4\\2\\4\\2\end{bmatrix}, \quad \begin{bmatrix}8\\8\\4\\8\\4\end{bmatrix}, \quad \begin{bmatrix}10\\4\\8\\4\\8\end{bmatrix}, \quad \begin{bmatrix}10\\8\\6\\8\\6\end{bmatrix}, \quad \begin{bmatrix}9\\6\\6\\6\\6\end{bmatrix}$$

You are asked to identify some representative customers (in terms of their tastes in recipes), so that each customer's tastes can be modeled in terms of these representatives.

a) Find a basis for the customer tastes.

SOLUTION: Denote the columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_5$. First note that first two columns are linearly dependent, but the second and third are not. These form the first two vectors in our basis. Now we need to see whether the fourth and fifth are linearly dependent on the second and third. To see how to do this, we try to find numbers α and β such that $\boldsymbol{x}_4 = \alpha \boldsymbol{x}_2 + \beta \boldsymbol{x}_3$. We have a system of 5 equations:

$$\begin{aligned} 8\alpha + 10\beta =& 10 \\ 8\alpha + 4\beta =& 8 \\ 4\alpha + 8\beta =& 6 \\ 8\alpha + 4\beta =& 8 \\ 4\alpha + 8\beta =& 6 \end{aligned}$$

The first equation implies that $\alpha = (5/4)(1 - \beta)$. The second equation implies $\alpha = (1/2)(2 - \beta)$. Together these imply $\beta = 1/3$ and $\alpha = 5/6$. Now we need to see whether these values satisfy the remaining three equations. They do, so $\boldsymbol{x}_4$ is linearly dependent on $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. (Note we only needed to check the third equation because the fourth equation is the same as the second and the fifth equation is the same as the third.) We repeat this exercise for $\boldsymbol{x}_5$; it holds for $\alpha = \beta = 1/2$. Thus we can form a basis as

$$\begin{bmatrix}8\\8\\4\\8\\4\end{bmatrix}, \quad \begin{bmatrix}10\\4\\8\\4\\8\end{bmatrix}$$

b) A new customer joins who rates Beef stew 5 and Carrot salad 7. (i) What do you predict will be their rating for Eggs benedict? (ii) Could you have figured this out if we only knew their rating for Carrot Salad? (iii) Could you have figured this out if we only knew rating for Beef stew?

SOLUTION: (i) Easy way: note equations 4-5 are the same as equations 2-3, so we immediately know that the Eggs benedict rating is always equal to the Carrot salad rating – in this case, 7.

Hard way: this new user is modeled as a weighted sum of our two basis vectors. So we need to find α and β so that

$$\begin{aligned} 8\alpha + 4\beta &= 5 \\ 4\alpha + 8\beta &= 7 \end{aligned}$$

This holds for $\alpha = 1/4$ and $\beta = 3/4$. From here we can determine that the predicted Eggs benedict rating is $4\alpha + 8\beta = 7$.

(ii) Yes, we could have, because Egg rating is always equal to Carrot rating.

(iii) No, there's no enough information to make this determination.

5. Suppose that a fish packing plant wants to automate the process of sorting incoming fish on a conveyor belt according to species. You decide to try to separate sea bass from salmon using optical sensing. You set up a camera, take some sample images and begin to note some physical differences between the two types of fish, such as length and lightness.

a) You first try to build a linear classifier using least squares, with a decision rule of

$$\tilde{y} = \begin{cases} +1 & \text{if } w_1 \cdot \text{length } + w_2 \cdot \text{lightness } > t \\ -1 & \text{otherwise} \end{cases}.$$

Describe (1) how you would formulate this as a least-squares problem to find w_1, w_2, and t that best fit the training data you've gathered and (2) how you would solve this least squares problem.

SOLUTION: Equivalent rule is

$$\tilde{y} = \begin{cases} +1 & \text{if } w_1 \cdot \text{length } + w_2 \cdot \text{lightness } + w_3 > 0 \\ -1 & \text{otherwise} \end{cases}$$

where $w_3 = -t$. Let

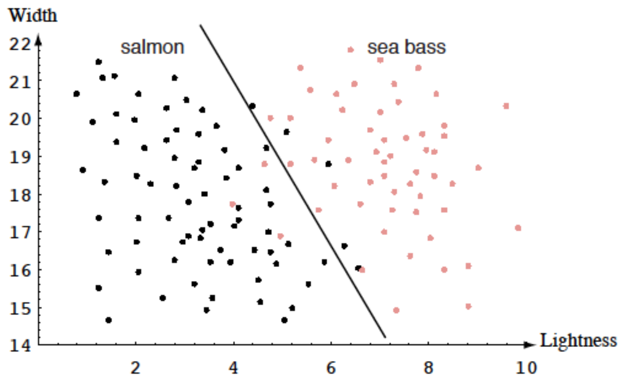
$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^\top \qquad \text{and} \qquad X = \begin{bmatrix} \text{length}_1 & \text{lightness}_1 & 1 \\ \text{length}_2 & \text{lightness}_2 & 1 \\ \vdots & \vdots & \vdots \\ \text{length}_n & \text{lightness}_n & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

If columns of X are linearly independent, then

$$\widehat{\boldsymbol{w}} = (X^\top X)^{-1} X^\top \boldsymbol{y}$$

and I apply the decision rule above.

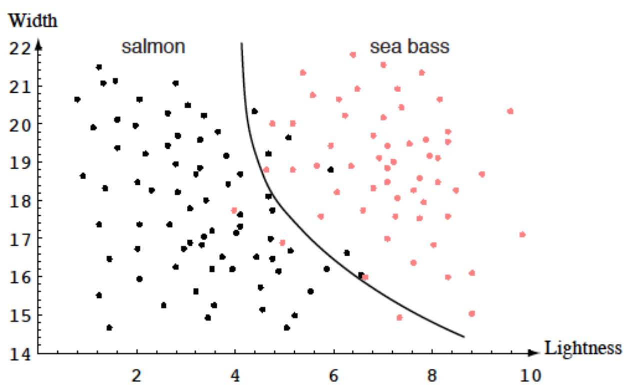
b) After learning your linear least squares classifier, you plot the decision boundary and see this:



You then decide to see whether you can build a classifier with a *quadratic* decision boundary. That is, the decision boundary is the set of (length,lightness) pairs that satisfy

$$w_1 + w_2 \cdot \text{length} + w_3 \cdot \text{lightness} + w_4 \cdot \text{length}^2 + w_5 \cdot \text{lightness}^2 + w_6 \cdot \text{length} \cdot \text{lightness} = 0$$

for some set of coefficients $w_1, \ldots, w_6$, as in this image:



Describe a method for learning this classifier from the same data as you had with the linear classifier above.

SOLUTION: Let

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^\top \qquad \text{and} \qquad \boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix}$$

and

$$X = \begin{bmatrix} 1 & \text{length}_1 & \text{lightness}_1 & \text{length}_1^2 & \text{lightness}_1^2 & \text{length}_1 \cdot \text{lightness}_1 \\ 1 & \text{length}_2 & \text{lightness}_2 & \text{length}_2^2 & \text{lightness}_2^2 & \text{length}_2 \cdot \text{lightness}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{length}_n & \text{lightness}_n & \text{length}_n^2 & \text{lightness}_n^2 & \text{length}_n \cdot \text{lightness}_n \end{bmatrix}.$$

If columns of X are linearly independent, then

$$\widehat{\boldsymbol{w}} = (X^\top X)^{-1} X^\top \boldsymbol{y}$$

and I apply the decision rule

$$\tilde{y} = \begin{cases} +1 & \text{if } \widehat{w}_1 + \widehat{w}_2 \cdot \text{length} + \widehat{w}_3 \cdot \text{lightness} + \widehat{w}_4 \cdot \text{length}^2 + \widehat{w}_5 \cdot \text{lightness}^2 + \widehat{w}_6 \cdot \text{length} \cdot \text{lightness} > 0 \\ -1 & \text{otherwise} \end{cases}.$$