STAT 489
Spring 2020
Instructor: Minh Pham
Final Exam
05/05/2020
Due date May 6th, 8:00 AM, Eastern Time.

On my honor as a student I have neither given nor received aid on this examination.

Name (Print): Owen McClure

Email ID (Print): odm7341

Signature: Owen McClure

This exam booklet contains 7 pages (including this cover page) and 6 problems. Check to see if any pages are missing. Enter all requested information on the top of this page, and put your initials on the top of every page, in case the pages become separated.

Please only hand in your exam booklet after exam

You may *not* use any external source other than the allowed written notes on this exam.

You are required to show your work on each problem on this exam. The following rules apply:

- **Organize your work**, in a reasonably neat and coherent way, in the space provided. If you need more space, use the back of the pages; clearly indicate when you have done this, otherwise you will not receive the credit.
- **Unsupported answers will not receive full credit.** Unless noted otherwise, a correct answer, unsupported by calculations, derivations or explanation, will receive no credit; an incorrect answer supported by substantially correct work might still receive partial credit.
- **Avoid multiple solutions..** Please only provide one solution to each problem. If multiple solutions are presented, the worst solution will be used to determine your credit.
- **Only communicate with the instructor about the exam. You are not allowed to discuss it with your classmates.**

Problem	Points	Score
1	10	
2	16	
3	10	
4	10	
5	10	
6	10	
Total:	66	

Do not write in the table to the right.

1. (10 points) The ridge regression problem is given by solving the optimization problem:

$$\beta^{\hat{ridge}} = min_{\beta}\left\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right\},$$

where β_0 is the intercept of the regression model needed to be estimated, β_j's are the corresponding coefficients for feature j. Let $\bar{x}_j$ be the mean (average) of feature j (Average of $x_{1j}, x_{2j}, \cdots, x_{Nj}$). Show that the ridge problem is equivalent to the following problem (i.e if you know the solution of one problem, you will know the solution of the other problem):

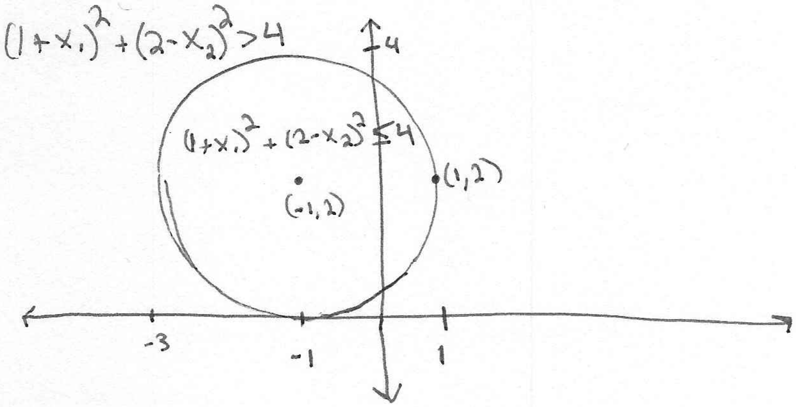
$$\hat{\beta^c} = min_{\beta^c}\left\{\sum_{i=1}^{N}(y_i - \beta_0^c - \sum_{j=1}^{p}(x_{ij} - \bar{x}_j)\beta_j^c)^2 + \lambda\sum_{j=1}^{p}\beta_j^{c2}\right\}.$$

The second problem is simply ridge for centered X. It can be rewritten with $\beta_0^c = \beta_0 + \beta_j \bar{X}_j$ and then becomes $\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$ which is the same as the given ridge problem. This is because β_j only depends on the variance of X_j and not its mean. The only difference between both of these solutions is the constant's value.

2. (16 points) Consider a classification problem with $p = 2$ predictors and a non-linear decision boundary:

   1. Sketch the curve:(4pts)

$$(1 + x_1)^2 + (2 - x_2)^2 = 4.$$



   2. On your sketch indicate the region for which: (4pts)

$$(1 + x_1)^2 + (2 - x_2)^2 > 4.$$ = outside the circle, not including the boundary

as well as:

$$(1 + x_1)^2 + (2 - x_2)^2 \leq 4.$$ = inside the circle, including the boundary

   3. Suppose a classifier assigns an observation to class 1 if $(1 + x_1)^2 + (2 - x_2)^2 > 4$, class -1 if $(1 + x_1)^2 + (2 - x_2)^2 \leq 4$. To what class is the observation (0,0), (2,2) classified?(4pts)

(0,0) and (2,2) are classified as 1

   4. Show that although the classifider is not linear in terms of x_1 and x_2 but it is linear interms of x_1,x_1^2,x_2,x_2^2.(4pts)

This can be shown by rewriting the equation

$(1+x_1)^2+(2-x_2)^2 \leq 4$

$x_1^2+x_2^2+2x_1-4x_2+5 \leq 4$

Now this equation is linear in terms of x_1, x_2, x_1^2, x_2^2

3. (10 points) A more generalized lasso penalty for $\beta \in \mathcal{R}^p$ is defined as:

$$Lasso(\beta) = \lambda_1|\beta_1| + \lambda_2|\beta_2| + \cdots + \lambda_p|\beta_p|.$$

for $\lambda_i > 0$. In this problem, you are going to show that this lasso penalty is a convex function.

1. Prove that each component of this penalty is a convex function, i.e: $\lambda_i|\beta_i|$ is a convex function. (6pts) (Hint: This is convex function, not convex set).

consider the function $f(\beta) = \lambda|\beta|$

and $\forall x, y \in \mathbb{R}$, $\forall a, b \geq 0$ where $a+b=1$

$$f(ax+by) = \lambda|ax+by|$$

and by the triangle inequality $\lambda|ax+by| \leq \lambda(|ax|+|by|)$

$= \lambda(|a||x| + |b||y|) = a\lambda|x| + b\lambda|y| = af(x) + bf(y)$

Therefore $f(ax+by) \leq af(x) + bf(y)$ which is the definition of convexity ∎

2. Prove that the whole penalty function is a convex function. (4pts)

Consider any 2 convex functions $f(x)$ & $g(x)$ and $h(x) = f(x) + g(x)$

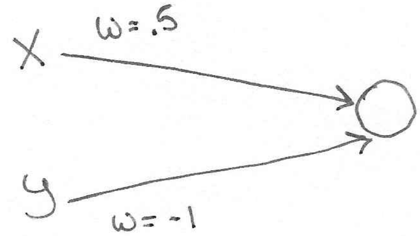
Then $\forall x, y \in \mathbb{R}$, $\forall a, b \geq 0$ where $a+b=1$

$$f(ax+by) + g(ax+by) \leq af(x) + bf(y) + ag(x) + bg(y)$$
$$h(ax+by) \leq a(f(x)+g(x)) + b(f(y)+g(y))$$
$$h(ax+by) \leq a\,h(x) + b\,h(y)$$

This proves the sum of any 2 convex functions is also convex. Since the penalty function is just a sum of convex functions, it is also convex ∎

4. (10 points) Consider the following data set with 4 data points: (1,1), (1,0), (0,0), (0,1) with corresponding labels: 0, 1, 0, 0. Can you use a single perceptron unit with threshold function to classify this data set? If yes, show the weights. (The threshold function g(x) return 1 if $x \geq 0$, 0 otherwise.)

yes



5. (10 points) Consider the regression problem with real-valued data Y and X. Y is generated conditional on X from the following process:

$$\epsilon \sim N(0, \sigma^2), Y = aX + \epsilon.$$

ϵ is the noise term drawn from a Gaussian distribution. Assume we have training data set of n pairs $(X_i, Y_i), i = 1, 2, \cdots, n$. To estimate a we solve the following optimization problem:

$$\min_a \frac{1}{2}\sum_{i=1}^{n}(Y_i - aX_i)^2$$

Solve for the estimate for a from the training data above.

$$= \frac{d}{da}\left(\min_a \frac{1}{2}\sum_{i=1}^{n}(y_i - ax_i)^2\right)$$

$$= \sum (y_i - a x_i)(-x_i) = 0$$

$$= -\sum x_i y_i + \sum a_i x_i^2 = 0$$

$$= a_i \sum x_i^2 = \sum x_i y_i$$

$$a_i = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

6. (10 points) Given the design matrix X= $\begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix}$

Use PCA to reduce the dimension from 2 to 1. Hints: Follow the steps in the lecture: center the data, compute covariance matrix, calculate eigen values and eigen vectors...

$$\frac{6-3-2+7}{4} = 2 \qquad \frac{(6)^2+(-3)^2+(-2)^2+(7)^2}{3} = 32.67$$

$$\sigma_{x_1} = 5.72$$

$$\frac{-4+5+6-3}{4} = 1 \qquad \frac{(-4)^2+5^2+6^2+(-3)^2}{3} = 28.07$$

$$\sigma_{x_2} = 5.35$$

$$X_{norm} = \begin{bmatrix} .699 & -.935 \\ -.874 & .748 \\ -.699 & .935 \\ .874 & -.748 \end{bmatrix} \qquad \frac{1}{3} X^T X = \frac{1}{3} \begin{bmatrix} 2.505 & -2.615 \\ -2.615 & 2.867 \end{bmatrix}$$

$$= \begin{bmatrix} .835 & -.872 \\ -.872 & .956 \end{bmatrix}$$

$$X^T_{norm} = \begin{bmatrix} .699 & -.874 & -.699 & .874 \\ -.935 & .748 & .935 & -.748 \end{bmatrix}$$

$$\lambda_1 = 0.02 \quad V_1 = \begin{bmatrix} 1.07 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 1.77 \quad V_2 = \begin{bmatrix} -0.93 \\ 1 \end{bmatrix}$$

$$\boxed{(X_{norm}) V_2 = \begin{bmatrix} -1.585 \\ 1.561 \\ 1.585 \\ -1.561 \end{bmatrix}}$$