Instructions

1. It is a CLOSED BOOK examination.
2. This paper has three questions. The maximum marks is 20.
3. Write your answers on a paper, scan and submit them at the end of the exam.
4. Write your name, roll number and the subject number (CS 419M) on the top of each of your answer script.
5. There will be partial credits for subjective questions, if you have made substantial progress towards the answer. However there will be NO credit for rough work.
6. Please keep your answer sheets different from the rough work you have made. Do not attach the rough work with the answer sheet. You should ONLY upload the answer sheets.

1. Consider a dataset of 1-D points, $\mathcal{D} = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\}$, $x_i \in \mathbb{R}, x_i > 0, y_i \in \mathbb{R}$. We want to find a lasso regression estimate w_R such that:

$$w_R = \arg\min_{w'} \left(\sum_{i=1}^{N} (y_i - w'x_i)^2 + \lambda w \right)$$

Note that in above expression summation is over the whole equation $(y_i - w'x_i)^2 + \lambda w$

Here, $\lambda \geq 0$ is the lasso regression coefficient. Each y_i is assumed to be generated such that $y_i = wx_i + \varepsilon_i$ where $\varepsilon_i \sim \mathrm{N}(0, 1)$ (i.e. ε_i is zero-mean unit-variance Gaussian noise) and w is the true (unknown) linear relationship that we would like to estimate.

1.a Derive a closed-form expression for w_R in terms of α, β, N and λ, where

$$\alpha = \sum_{i=1}^{N} x_i^2, \quad \beta = \sum_{i=1}^{N} x_i y_i$$

1.a	/2	

Solution:-

$$w' = \frac{2\beta - \mathrm{N}\lambda}{2\alpha}$$

1.b What is $\mathrm{E}[w_R]$ where the expectation is taken with respect to all y_i 's? Write down your result in terms of w, α, β, N and λ.

1.b	/2	

Solution:-

$$\mathrm{E}[w_R] = \frac{2w\alpha - \mathrm{N}\lambda}{2\alpha}$$

1.c $\hat{\theta}$ is said to be an unbiased estimator of the true parameter θ if $E[\hat{\theta}] = \theta$. For what value of λ will w_R be an unbiased estimator of w ? Justify your answer.

1.c	/1	

Solution:- $\lambda = 0$

2. Consider the following objective function $L(\mathbf{w})$ parameterized by a weight vector $\mathrm{w} \in \mathbb{R}^2$:

$$L(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathbf{U} \mathbf{w}$$

where

$$\mathbf{U} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

Say we want to minimize this loss function using gradient descent. If w^t is the weight vector after t iterations of gradient descent, write down the gradient descent update equation for $\mathbf{w}^{t+1}$ using a learning rate of $\eta = \frac{1}{5}$. Reduce this expression so that w^{t+1} is in the form $\mathbf{w}^{t+1} = \mathbf{V}\mathbf{w}^t$ where $\mathbf{V}$ is a matrix.

2.a Calculate matrix $\mathbf{V}$?

2.a	/1	

Solution:

$$\begin{aligned}
\mathbf{w}^{t+1} &= \mathbf{w}^t - \eta \nabla L\left(\mathbf{w}^t\right) \\
\nabla L\left(\mathbf{w}^t\right) &= \mathbf{U}\mathbf{w}^t \\
\mathbf{w}^{t+1} &= (I - \eta \mathbf{U})\mathbf{w}^t \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \right) \mathbf{w}^t \\
&= \left(\begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{3}{5} \end{bmatrix} \right) \mathbf{w}^t
\end{aligned}$$

$$\mathbf{V} = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{3}{5} \end{bmatrix}$$

2.b $L(\mathbf{w})$ is minimized at $\mathbf{w}^* = 0$. Show that we can converge to the optimal $\mathbf{w}^*$ using gradient descent starting from any arbitrary initialization for the weight vector w_0.

2.b	/1	

Solution: The optimal $w^* = 0$ and $\mathbf{V}$ is diagonal, ensuring that in the limit (i.e., $t \to \infty$, the non-zero diagonal values will converge to 0 .

2.c Instead of $\mathbf{U}$, say we use a diagonal matrix $\mathbf{Z} = \begin{pmatrix} z_{11} & 0 \\ 0 & z_{22} \end{pmatrix}$ where z_{11} and z_{22} are both greater than 0 . What value of η, in terms of z_{11} and z_{22}, would lead to fastest convergence?

2.c	/3	

Solution: The diagonal entries after the weight update rule will be $|1-\eta z_{11}|$ and $|1-\eta z_{22}|$. We want to minimize $\max\left(|1-\eta z_{11}|, |1-\eta z_{22}|\right)$. Let $\alpha = \min\left(z_{11}, z_{22}\right)$ and $\beta = \max\left(z_{11}, z_{22}\right)$. Then, $\max\left(|1-\eta z_{11}|, |1-\eta z_{22}|\right) = \max(|1-\eta\alpha|, |1-\eta\beta|)$, where $\alpha \leq \beta$ Now,

$$(|1-\eta\alpha|, |1-\eta\beta|) = \begin{cases} (1-\eta\alpha, 1-\eta\beta) & \text{if } \eta \leq \frac{1}{\beta} \\ (1-\eta\alpha, \eta\beta-1) & \text{if } \frac{1}{\beta} \leq \eta \leq \frac{1}{\alpha} \\ (\eta\alpha-1, \eta\beta-1) & \text{if } \eta \geq \frac{1}{\alpha} \end{cases}$$

Also, note that $1-\eta\alpha > 1-\eta\beta$, and $\eta\beta-1 > \eta\alpha-1$. Hence we have

$$\max(|1-\eta\alpha|, |1-\eta\beta|) = \begin{cases} 1-\eta\alpha & \text{if } \eta \leq \frac{1}{\beta} \\ \max(1-\eta\alpha, \eta\beta-1) & \text{if } \frac{1}{\beta} \leq \eta \leq \frac{1}{\alpha} \\ \eta\beta-1 & \text{if } \eta \geq \frac{1}{\alpha} \end{cases}$$

Let η_0 be such that $1-\eta_0\alpha = \eta_0\beta-1$: i.e., $\eta_0 = 2/(\alpha+\beta)$. Then for $\eta \in \left[\frac{1}{\beta}, \eta_0\right]$ we have $1-\eta\alpha \geq \eta\beta-1$ and for $\eta \in \left[\eta_0, \frac{1}{\alpha}\right]$ we have $\eta\beta-1 \geq 1-\eta\alpha$. Hence, we have

$$\max(|1-\eta\alpha|, |1-\eta\beta|) = \begin{cases} 1-\eta\alpha & \text{if } \eta \leq \eta_0 \\ \eta\beta-1 & \text{if } \eta \geq \eta_0 \end{cases}$$

Finally, we note that $1-\eta\alpha$ is a decreasing function, and $\eta\beta-1$ is an increasing function of η. Hence the above function is minimized at $\eta = \eta_0 = 2/\left(z_{11}+z_{22}\right)$ (where it attains the value $\frac{\beta-\alpha}{\beta+\alpha} = \frac{|z_{11}-z_{22}|}{|z_{11}+z_{22}|}$).

Total: 10