Consider the following maximum likelihood estimation(MLE) objective for linear regression:

$$\max_{\theta} \prod_{i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right)$$

which leads to the following objective (by taking -ve log):

$$\min_{\theta} \sum_{i} (y_i - \theta^T x_i)^2$$

In the MLE objective,

(A) **[Ans]** It is assumed that the residuals $(y_i - \theta^T x_i)$ have a mean of 0

(B) It is assumed that the residuals $(y_i - \theta^T x_i)$ have a standard deviation of 1

(C) **[Ans]** It is assumed that all the residuals $(y_i - \theta^T x_i)$ have the same standard deviation

(D) None of these

We saw the loss function for linear regression as

$$J(\theta) = (Y - X\theta)^T (Y - X\theta)$$

We saw that we get a closed form solution for θ by solving $\frac{\partial J(\theta)}{\partial \theta} = 0$:

$$\frac{\partial}{\partial \theta} \left(Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta \right) = 0$$

$$\implies -2X^T Y + 2X^T X\theta = 0 \implies \theta = (X^T X)^{-1} X^T Y$$

Now find the closed form solution that minimizes this loss function (assume W is symmetric):

$$J(\theta) = (Y - X\theta)^T W (Y - X\theta)$$

(A) $\theta = (X^T W X)^{-1} X^T Y$

(B) $\theta = (X^T X)^{-1} X^T W Y$

(C) **[Ans]** $\theta = (X^T W X)^{-1} X^T W Y$

(D) $\theta = (X^T W X)^{-1} X^T W^{-1} Y$

(E) None of these

Consider the function

$$f(w) = w^2 + w + 1$$

We want to find the minima of the function using gradient descent. We start at $w^0 = 5$. What should be the learning rate η so that we reach the minima in a single step?

Hint: There may be many ways to solve this. One of the easiest is to see that the derivative of the point after 1st update is 0:

(A) 1

(B) **[Ans]** 0.5

(C) 0.1

(D) 0.05

(E) None of these

Let us say that we have computed the gradient of our cost function and stored it in a vector g. What is the cost of one gradient descent update given the gradient?

D is number of dimensions, N is the number of samples

(A) **[Ans]** $O(D)$

(B) $O(N)$

(C) $O(ND)$

(D) $O(ND^2)$

Consider the dataset of 4 points in R^2

$$X = \begin{bmatrix} 7 & -3 \\ 6 & -4 \\ -2 & 6 \\ -3 & 5 \end{bmatrix}$$

Run PCA for this data (in your notebook) to go from R^2 to R^1. Then

(A) 1st Principal component is $[1, 1]^T$

(B) **[Ans]** 1st Principal component is $[1, -1]^T$

(C) **[Ans]** The projection of 1st and 2nd points in the new subspace is at same point

(D) **[Ans]** The projection of 3rd and 4th points in the new subspace is at ame point

(E) The projection of 1st and 3rd points in the new subspace is at same point

(F) None of these