Quiz 1

Wednesday, February 19, 2025 4:40 PM

1. A prompt to an AI assistant to implement linear regression using gradient descent produces the following expression for the gradient vector. Is this equivalent to the expression that was derived in class? Justify your answer.

$$\nabla = X^T(X\underline{w} - \underline{t})$$

2. You are given N data points and are told that they were generated by an exponential distribution (below). (a) Derive an expression that uses the data to find an estimate of the parameter of this distribution. (b) What assumption on the data must hold?

$$p(x;\lambda) = \lambda \exp(-\lambda x) \quad x \geq 0$$

3. A model is found to be underfitting. Should the strength of the regularization term in the loss function be decreased or increased? Why?
4. Suppose X and Y are independent random variables. Show that $E(X|Y) = E(X)$.

7 pts each
Max: 30 pts
Min: 2 pts

1. $X^T(X\underline{w} - \underline{t}) = X^TX\underline{w} - X^T\underline{t}$ (3)

take the transpose of this expression → $(X^TX\underline{w} - X^T\underline{t})^T = \underline{w}^TX^TX - \underline{t}^TX$ (3)

this is proportional to the gradient derived in class ⇒ Yes, they are equivalent (1)

2. the likelihood $\prod_{i=1}^{N} \lambda e^{-\lambda x_i}$ (2)

log-likelihood $\log \prod_{i=1}^{N} \lambda e^{-\lambda x_i} = \sum_{i=1}^{N} \log\lambda - \lambda x_i$

$= N(\log\lambda - \sum_{i=1}^{N} \lambda x_i$ (2)

Maximize

$\frac{\partial(\cdot)}{\partial\lambda} = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i = 0$ (2) $\Rightarrow \lambda = \frac{N}{\sum_{i=1}^{N} x_i}$ (1)

3. Decrease. Model Complexity needs to be increased. (7)

4. $E(X|Y) = \int x\, p(x|y)\, dx$ (2)

since X & Y are S.I.

(3) $p(x,y) = p(x|y)\,p(y) = p(x)\,p(y)$

$= \int x\, p(x)\, dx$ (2) $= E(X)$