CS 6140: Machine Learning — Fall 2021— Paul Hand

HW 3

Due: Wednesday October 6, 2021 at 2:30 PM Eastern time via Gradescope.

Names: [Put Your Name(s) Here]

You can submit this homework either by yourself or in a group of 2. You may consult any and all resources. Make sure to justify your answers. If you are working alone, you may either write your responses in LaTeX or you may write them by hand and take a photograph of them. If you are working in a group of 2, you must type your responses in LaTeX. You are encouraged to use Overleaf. Create a new project and replace the tex code with the tex file of this document, which you can find on the course website. To share the document with your partner, click Share > Turn on link sharing, and send the link to your partner. When you upload your solutions to Gradescope, make sure to take each problem with the correct page or image.

Question 1. *Linear regression with multivariate responses.*

Consider training data $\{(x^{(i)}, y^{(i)})\}_{i=1\ldots n}$, where $x^{(i)} \in \mathbb{R}^d$ and $y^{(i)} \in \mathbb{R}^k$. Consider a model $y = Ax$, where $A \in \mathbb{R}^{k\times d}$ is unknown. Estimate A by solving least squares linear regression

$$\min_A \sum_{i=1}^{n} \|y^{(i)} - Ax^{(i)}\|^2.$$

(a) Find A in the case of training data $\left\{\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}\right), \left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right), \left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2\\3\\1\end{pmatrix}\right)\right\}$. You may use a computer to perform linear algebra. Hint: the problem can be simplified by observing that each output dimension can be computed separately from the others. If you use this fact, justify it in your response.

Response:

(b) Consider the case of generic training data. Let Y be the $k \times n$ matrix such that $Y_{ji} = y_j^{(i)}$. Let X be the $n \times d$ matrix where $X_{ij} = x_j^{(i)}$. Provide a formula for the least squares estimate of A. Make sure to check that the matrix dimensions match in any matrix products that appear in your answer. Use the same hint as in part (a).

Response:

(c) Show that any prediction under this learned model is a linear combination of the response values $(y^{(1)}, \ldots, y^{(n)})$. That is, for the A in part (b), show that $Ax \in \text{span}(y^{(1)}, \ldots, y^{(n)})$ for any x. You may assume that X is rank d.

Response:

Question 2. *Logistic Regression*

Consider training data $\{(x_i, y_i)\}_{i=1\ldots n}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{0,1\}$. Consider the logistic data model $\hat{y} = \sigma(\theta \cdot x)$, where $x \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$, and σ is the logistic function $\sigma(z) = e^z/(e^z + 1)$.

(a) Show that $\sigma'(z) = \sigma(z)(1 - \sigma(z))$.

Response:

(b) Let $f(\theta) = \sum_{i=1}^{n} -y_i \log \hat{y}_i - (1 - y_i)\log(1 - \hat{y}_i)$, where $\hat{y}_i = \sigma(\theta \cdot x_i)$. Compute $\nabla f(\theta)$. Use the fact in part (a) to simplify your answer.

Response:

(c) If $M = \sum_{i=1}^{n} x_i x_i^t$, show that $z^t M z \geq 0$ for any $z \in \mathbb{R}^d$.

Response:

(d) Using a summation and vector and/or matrix products, write down a formula for the Hessian, H, of f with respect to θ. Show that $z^t H z \geq 0$ for any $z \in \mathbb{R}^d$.

Response: