DSC 140A - Discussion 01

Welcome to Discussion 01. This week's discussion (and most of the future discussions) will be focused around a worksheet. Some problems on this worksheet are designed to be useful for getting started with this week's homework. Other problems are meant to remind you of concepts from prerequisite courses that will be useful for this week's lectures.

During discussion, we'll give you time to work on these problems in small groups. During this time, the TA will be walking around to help you with any questions you might have – raise your hand if you need help. After a few minutes, we'll reconvene as a class to go over the solution.

You're encouraged to look at the problems and give them a try before discussion section, but it's not required.

Problem 1.

Let $\vec{w} \in \mathbb{R}^d$, and define $f(\vec{w}) = \|\vec{w}\|^2$.

In this problem, we'll show that $\frac{df}{d\vec{w}} = 2\vec{w}$. That is, the *gradient* of f is $2\vec{w}$.

The general process for finding the gradient is:

1. "Expand" f so that you see all of the coordinates of $\vec{w}$. That is, $w_0, w_1, \ldots, w_d$.
2. "Differentiate" by taking partial derivatives with respect to w_0, w_1, ..., w_d.
3. "Recombine" the partial derivatives into a vector (a.k.a., the *gradient* vector).

Note: you'll follow a similar process on the homework to show that the the gradient of $f(\vec{x}) = x^T Ax$ is $2A\vec{x}$.

a) Show that $f(\vec{w}) = \sum_{i=1}^{d} w_i^2$.

b) Show that $\frac{\partial f}{\partial w_i} = 2w_i$.

c) In vector form, what is $\frac{df}{d\vec{w}}$?

Problem 2.

Let A be a *symmetric* $n \times n$ matrix, with entries:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

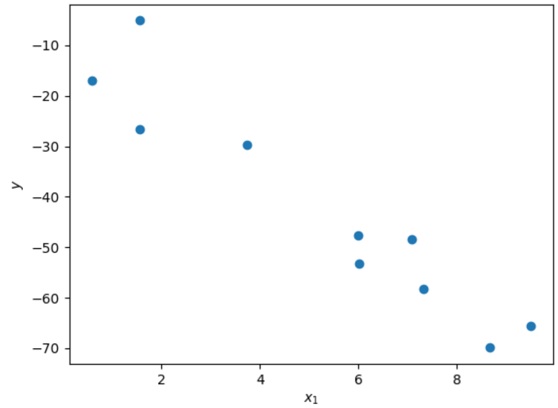
Suppose $\vec{x} \in \mathbb{R}^n$.

Show that $\vec{x}^T A\vec{x} = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij}$.

Hint: this might be useful for the homework.

Problem 3.

Consider the data shown below:



Let $R(\vec{w})$ be the risk with respect to the mean squared error for this data, assuming a linear prediction function of the form $H(x) = w_0 + w_1 x$.

Let $\vec{g}$ be the gradient of $R(\vec{w})$ evaluated at the point $(40, 20)$. In other words, let $\vec{g} = \frac{dR}{d\vec{w}}(40, 20)$.

What will be the sign (positive, negative, or zero) of the first component of $\vec{g}$? What about the second component?