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 $f(\vec{x}) = x^T A x$ is $2A\vec{x}$.

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

Solution:

$$\begin{aligned} f(\vec{w}) &= \|\vec{w}\|^2 \\ &= \vec{w} \cdot \vec{w} \\ &= \sum_{i=1}^d w_i^2 \end{aligned}$$

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

Solution:

$$\begin{aligned} \frac{\partial f}{\partial w_i} &= \frac{\partial}{\partial w_i}\left(\sum_{j=1}^d w_j^2\right) \\ &= \sum_{j=1}^d \frac{\partial}{\partial w_i}\left(w_j^2\right) \end{aligned}$$

This derivative will be zero when $i \neq j$, and $2w_i$ when $i = j$. This has the effect of picking out only the ith term in the sum (the other terms are zero).

$$= \underbrace{0}_{j=0} + \underbrace{0}_{j=1} + \cdots + \underbrace{2w_i}_{j=i} + \cdots + \underbrace{0}_{j=d}$$
$$= 2w_i$$

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

Solution: Since the gradient is a vector whose components are the partial derivatives of the function with respect to each of the variables, we have:

$$\begin{aligned}\frac{df}{d\vec{w}} &= \begin{pmatrix} \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_d} \end{pmatrix} \\ &= \begin{pmatrix} 2w_1 \\ \vdots \\ 2w_d \end{pmatrix} \\ &= 2 \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} \\ &= 2\vec{w}\end{aligned}$$

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

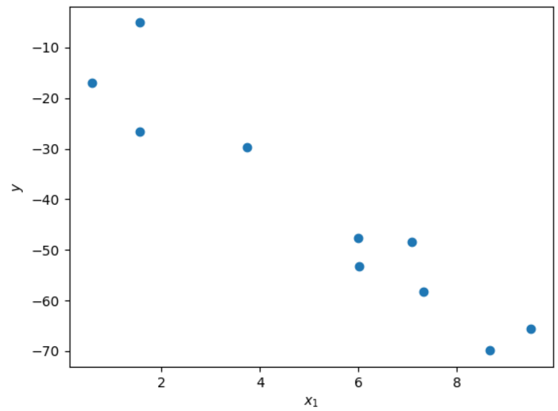
Hint: this might be useful for the homework.

Solution:

$$\begin{aligned}\vec{x}^T A \vec{x} &= \vec{x}^T \sum_{j=1}^{n} x_j \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \\ &= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j a_{ij} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}\end{aligned}$$

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?

Solution: The gradient points in the direction of steepest ascent. Therefore, this question is essentially asking: in order to increase the risk, should we increase or decrease w_0? What about w_1?

Glancing at the data, the line of best fit will have an intercept (w_0) of near zero and a negative slope (w_1).

The current point in question is $(w_0, w_1) = (40, 20)$. This corresponds to a line whose intercept is far too positive, and whose slope is far too positive. Therefore, to increase the risk even further, we should make w_0 even more positive (by increasing it) and make w_1 even more positive (by increasing it). So the signs of the components of $\vec{g}$ will both be positive.

Note: an earlier version of this problem asked about the gradient evaluated (-40 ,20), which was trickier. On one hand, an intercept of -40 is much too small for the data, and we expect the optimal choice to be close to zero. So it might seem that increasing w_0 will *decrease* risk, but this turns out not to be the case. Plotting the line with $w_0 = -40$ and $w_1 = 20$ will help see this: a small increase in w_0 will shift the prediction function upwards, causing some of the residuals to get smaller, but most of them to get larger, increasing risk. Thus the sign of the first entry of the gradient is still positive.