1 Derivatives

1.1 Sum of squared differences

Remember that the x_i and y_i are constant, and we're allowed to change the parameters a and b. If our cost function C is the SSD, then:

$$C(a,b) = \sum_i (y_i - (ax_i + b))^2. \tag{1}$$

Compute the derivatives of SSD with respect to a and b. *Hint: remember the chain rule if you don't want to multiply out the square.*

$$\frac{dC}{da} = \tag{2}$$

$$\frac{dC}{db} = \tag{3}$$

2 Setting up problems

2.1 Write linear regression with matrices

Write down this system of linear equations using matrix-vector multiplication. *Hint: Remember what size matrices are allowed to be left-or-right multiplied with each other*

$$\begin{aligned} y_1 &= \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 \\ y_2 &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \end{aligned} \tag{4}$$

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} & & \\ & ? & \\ & & \end{bmatrix} \tag{5}$$

2.2 Rewrite the same linear system

Rewrite your answer above so that y is a column vector.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} & & \\ & ? & \\ & & \end{bmatrix} \begin{bmatrix} & & \\ & ? & \\ & & \end{bmatrix} \tag{6}$$

3 Normal Equations

3.1 When do the Normal Equations not work ?

1. Write down a set of points x_i, y_i such that we can't use the normal equations to solve for the best-fit line. *Hint: When will $X^T X$ not be invertible?*

2. What results would PCA give for the points you gave above?

4 Challenges

4.1 Matrix "beta-squared" derivatives

Here we'll confirm that $\frac{\delta}{\delta \boldsymbol{\beta}}(\boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\beta}) = 2\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\beta}$

1. Show that $\boldsymbol{X}^T \boldsymbol{X}$ is a symmetric matrix. This means that if we rename it like this: $\boldsymbol{A} = \boldsymbol{X}^T \boldsymbol{X}$, then $a_{i,j} = a_{j,i}$

2. Write $\boldsymbol{\beta}^T \boldsymbol{A}$ as a linear combination of the rows of $\boldsymbol{A}$.

3. Write $\boldsymbol{\beta}^T \boldsymbol{A} \boldsymbol{\beta}$ as a linear combinations of dot products of rows of $\boldsymbol{A}$ and $\boldsymbol{\beta}$

4. Expand out the dot products as sums

5. Take the derivative with respect to one component of beta: β_i

6. Use the propery that $\boldsymbol{A}$ is symmetric

7. Concatenate into a vector, rearrange and QED! $\square$