DSC 40A - Reference Sheet for Final Part 1

Empirical Risk

For a loss function $L(h, y)$, the empirical risk is given by:

$$R(h) = \frac{1}{n}\sum_{i=1}^{n} L(h, y_i)$$

Simple Linear Regression

The parameters of the least squares regression line are:

$$w_1^* = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = r \cdot \frac{\sigma_y}{\sigma_x}$$

$$w_0^* = \bar{y} - w_1^*\bar{x}$$

Gradient Descent

The update rule is:

$$h_i = h_{i-1} - \alpha \cdot \left(\frac{dR}{dh}(h_{i-1})\right)$$

Convexity

A function $f : \mathbb{R} \to \mathbb{R}$ is convex if for every choice of a, b and $t \in [0, 1]$:

$$(1 - t)f(a) + tf(b) \geq f((1 - t)a + tb)$$

Linear Algebra for Regression

Least squares regression aims to minimize the mean squared error, which is:

$$R_{sq}(\vec{w}) = \frac{1}{n}||\vec{y} - X\vec{w}||^2$$

Minimizing this quantity is equivalent to satisfying the normal equations:

$$X^T X\vec{w} = X^T\vec{y}$$

If $X^T X$ is invertible, there is a unique solution to the normal equations, which is:

$$\vec{w}^* = (X^T X)^{-1}X^T\vec{y}$$