CS 480 Cheat Sheet

Perceptron

We assume $(x_1, y_1), ..., (x_n, y_n)$ belongs to some distribution.
Choose predictive function h_n such that $\max Pr(h(x_i) = y_i)$

Dot Product: $\langle w, x\rangle = \sum w_i x_i$
Padding: $\langle w, x\rangle + b = \langle (w, b), (x, 1)\rangle$
Note: $z = (w, b)$, $a_i = y_i(x_i, 1)$
Linear Seperable: if $s > 0$ and $Az > s1$

If data is not linearly seperable perceptron stalls.
Margin is determined by closest point to the hyperplane.
Perceptron finds a solution, no guarantee its the best solution.

l2 Norm: $||x||_2 = \sqrt{\sum_i x_i^2}$
Error Bound $\leq \frac{R^2||z||_2^2}{s^2}, R = \max ||a_i||_2$
Margin: $\gamma = \max_{||z||_2=1} \min_i \langle a_i, z\rangle$
One-versus-all: $\hat{y} = \text{argmax}_k \; w_k^T x + b_k$
One-versus-one: $\#\{k' : x^T w_{k,k'} + b_{k,k'} > 0 || x^T w_{k',k} + b_{k',k} < 0\}$

Example #1: classify $x = (1, 2)^T$ given $b = 1$ and $w = (2, -3)$:
$(1, 2, -3)(1, 1, 2)^T = -3 = -$
Example #2: given w=$(\frac{2}{3}, \frac{1}{3})$ and b = 0, whats the line:
$\frac{2}{3}x + \frac{1}{3}y = 0 \to y = -2x$

Linear Regression

Gradient if $f(x) : \mathbb{R}^d \to \mathbb{R}$, $\Delta f(v) = \left(\frac{\delta f}{\delta v_1}, ..., \frac{\delta f}{\delta v_d}\right) : \mathbb{R}^d \to \mathbb{R}^d$

Hessian $\Delta^2 f(v) = \begin{bmatrix} 1 & ... & 3 \\ ... & b & ... \\ 2 & ... & 3 \end{bmatrix}$