Linear Algebra

Basics

1. $w^T x_i = \sum_{j=1}^{d} w_j x_{ij}$, x_i, w is d x 1
2. $a^T Ab = b^T B^T a$ both sides are vectors
3. $\frac{1}{2}\|Xw - y\|_2^2 = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i) = \frac{1}{2}w^T X^T Xw - w^T X^T y + \frac{1}{2}y^T y$
4. $\nabla \text{const} = 0, \nabla w^t b = w, \nabla \frac{1}{2} w^T Aw = Aw$ if A symmetric
5. $\nabla \frac{1}{2}\|Xw - y\|_2^2 = X^T X - X^T y$
6. Normal equation $X^T Xw = X^T y$

Run Time

1. $X^T y : O(nd)$
2. $X^T X : O(nd^2)$
3. solve d x d system of equations : $O(d^3)$
4. solve normal equation $X^T y : O(d^3 + nd^2)$

Gradient Descent

1. $w^{t+1} = w^t - \alpha^t \nabla f(w^t) = w^t - X^T(Xw^t - y)$ (least square)
2. cost $O(nd)$ no need to form $X^T X$
3. total cost $O(ndt)$
4. faster for large d, works generally