Linear Regression

Notations

Input (features): $x \in R^d$

Output (target/label): $y \in R$

Data: $(x^{(1)}, y^{(1)}), ..., (x^{(N)}, y^{(N)})$

residual sum of squares (RSS): $RSS(w) = \Sigma_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})^2$

mean squared error (MSE): $MSE(w) = \frac{1}{N}RSS = E_{in}$

total sum of squares (TSS): $TSS = \Sigma_{i=1}^{N}(y^{(i)} - \bar{y})^2$

coefficient of determination = explained variation/total variation: $R^2 = 1 - \frac{RSS}{TSS}$

Simple Linear Regression

$J(w_0, w_1) = \frac{1}{2N}\sum_{i=1}^{N}((w_0 + w_1 x^{(i)}) - y^{(i)})^2$

Gradient Descent:

$\frac{\partial J(w_0,w_1)}{\partial w_0} = \frac{1}{N}\sum_1^N(w_0 + w_1 x^{(i)} - y^{(i)})$

$\frac{\partial J(w_0,w_1)}{\partial w_1} = \frac{1}{N}\sum_1^N(w_0 + w_1 x^{(i)} - y^{(i)})x^{(i)}$

for i = 1 to num_iter:

$temp0 = w_0 - \alpha\frac{\partial J(w_0,w_1)}{\partial w_0} = w_0 - \frac{\alpha}{N}\sum_1^N(w_0 + w_1 x^{(i)} - y^{(i)})$

$temp1 = w_1 - \alpha\frac{\partial J(w_0,w_1)}{\partial w_1} = w_1 - \frac{\alpha}{N}\sum_1^N(w_0 + w_1 x^{(i)} - y^{(i)})x^{(i)}$

$w_0 = temp0$

$w_1 = temp1$

Normal Equation Method:

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

Multiple Linear Regression

$J(w) = \frac{1}{2N}\sum_{i=1}^{N}((w^T x^{(i)}) - y^{(i)})^2$

Gradient Descent:

$\frac{\partial J(w)}{\partial w_0} = \frac{1}{N}\sum_1^N(w_0 x_0^{(i)} + w_1 x_1^{(i)} + ... + w_d x_d^{(i)} - y^{(i)})x_0^{(i)} = \frac{1}{N}\sum_1^N(w^T x^{(i)} - y^{(i)})x_0^{(i)}$ ($x_0^{(i)} = 1$
)

$$\frac{\partial J(w)}{\partial w_j} = \frac{1}{N}\sum_1^N (w_0 x_0^{(i)} + w_1 x_1^{(i)} + ... + w_d x_d^{(i)} - y^{(i)}) x_j^{(i)} = \frac{1}{N}\sum_1^N (w^T x^{(i)} - y^{(i)}) x_j^{(i)}$$

for i = 1 to num_iter:

$temp0 = w_0 - \alpha \frac{\partial J(w)}{\partial w_0} = w_0 - \frac{\alpha}{N}\sum_1^N (w^T x^{(i)} - y^{(i)}) x_0^{(i)}$

$temp1 = w_1 - \alpha \frac{\partial J(w)}{\partial w_1} = w_1 - \frac{\alpha}{N}\sum_1^N (w^T x^{(i)} - y^{(i)}) x_1^{(i)}$

...

$tempd = w_d - \alpha \frac{\partial J(w)}{\partial w_d} = w_d - \frac{\alpha}{N}\sum_1^N (w^T x^{(i)} - y^{(i)}) x_d^{(i)}$

$w_0 = temp0$

$w_1 = temp1$

...

$w_d = tempd$

Vector Implementation:

for i = 1 to num_iter:

$w = w - \alpha \nabla J(w) = w - \frac{\alpha}{N} X^T (Xw - y)$

Normal Equation Method:

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

Feature Scaling

min/max normalization: $x_j^{(i)} = \frac{x_j^{(i)} - \min(x_j)}{\max(x_j) - \min(x_j)}$

standardization: $x_j^{(i)} = \frac{x_j^{(i)} - ave(x_j)}{STD(x_j)}$