Problem Set 1
Solutions to Written Problems

1. Localized linear regression

Answer: First, re-write the cost as

$$E(x^{(q)}) = \frac{1}{2}\sum_{i=1}^{m}(y^{(i)} - \theta^T x^{(i)})^2 (w^{(i)})^2 = \frac{1}{2}\sum_{i=1}^{m}\left(w^{(i)}y^{(i)} - w^{(i)}\theta^T x^{(i)}\right)^2$$

Note that, unlike the SSD cost, this cost depends on the query example $x^{(q)}$. Let $\mathbf{W}$ be a diagonal m x m matrix with diagonal entries $w^{(i)}$, i = 1,...,m. Let us re-write the above cost equation in terms of matrices and vectors. Notice that multiplying the output vector $\mathbf{Y} = [y^{(1)} \quad ... \quad y^{(m)}]^T$ by $\mathbf{W}$ results in the vector of weighted outputs,

$$\mathbf{WY} = [w^{(1)}y^{(1)} \quad ... \quad w^{(m)}y^{(m)}]^T$$

and, similarly, multiplying the design matrix by $\mathbf{W}$ produces $\mathbf{WX}$, the weighted design matrix. Replacing the original $\mathbf{Y}$ and $\mathbf{X}$ in the SSD cost with their weighted versions, we can write the weighted cost function (showing vectors/matrices in boldface for clarity) as:

$$E(x^{(q)}) = \frac{1}{2}\sum_{i=1}^{m}\left(w^{(i)}y^{(i)} - w^{(i)}\theta^T x^{(i)}\right)^2 = (WY - WX\theta)^2$$

Substituting $\mathbf{Y} = \mathbf{WY}$ and $\mathbf{X} = \mathbf{WX}$ into the normal equations, the solution is

$$\theta_{ML} = ((WX)^T(WX))^{-1}(WX)^T(WY)$$

2. Maximum likelihood estimate for coin toss

Answer: The Maximum Likelihood solution is obtained by finding the value μ_{ML} that maximizes the likelihood function, which is the solution to

$$\frac{d}{d\mu}\left(\sum_{i=1}^{m} x^{(i)} ln\mu + (1 - x^{(i)}) ln(1-\mu)\right) = 0$$

Taking the derivative, we get

$$\sum_{i=1}^{m} x^{(i)}\frac{1}{\mu} + (-1)(1 - x^{(i)})\frac{1}{1-\mu} = 0,$$

$$\sum_{i=1}^{m} x^{(i)}\frac{1}{\mu} = \sum_{i=1}^{m}(1 - x^{(i)})\frac{1}{1-\mu},$$

$$\frac{h}{\mu} = \frac{m-h}{1-\mu}, \quad \mu = \frac{h}{m}$$

3. Maximum likelihood for logistic regression

Answer: First we write down the likelihood of the data given the parameters, which is

$$L(\theta) = \prod_{i=1}^{m} h(x^i)^{y^i}(1 - h(x^i))^{(1-y^i)}$$

Then we take log of both sides to get

$$lnL(\theta) = \sum_{i=1}^{m} h(x^i)y^i + (1 - h(x^i))(1 - y^i)$$

which is the same (up to a constant multiplier) as the logistic regression cost.