Machine Learning - Written Assignment 2

Sam Coxon

October 7, 2016

1. (a) $\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i) = \theta - \alpha \frac{1}{m} [(h_\theta(x_1) - y_1) + ... + (h_\theta(x_n) - y_n)]$

   (b) $h(x) = \theta^T \vec{x}$ where $\theta^T = [\theta_0, ..., \theta_n]$ and $\vec{x}^T = [1, x_1, ..., x_n]$

   (c) Gradient for $\theta_0 = \frac{1}{m}(\theta^T \vec{x} - y^T) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

   (d) Gradient for $\theta_i = \frac{1}{m}(\theta^T \vec{x} - y^T) \begin{bmatrix} x_i \\ x_i^2 \\ \vdots \\ x_i^m \end{bmatrix}$

   (e) $\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} := \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} - \alpha \frac{1}{m} [(\theta^T \vec{x} - y^T) * \begin{bmatrix} 1 & 1 & ... & 1 \\ x_1^0 & x_1^1 & ... & x_1^m \\ \vdots & \vdots & \vdots & \vdots \\ x_n^0 & x_n^1 & ... & x_n^m \end{bmatrix}]$

2. (a) Mean $= \mu = \frac{2+5+7+7+9+25}{6} = 9.17$

   Standard Dev. $= \sigma$

   $\sigma^2 = \frac{1}{m}[(2-9.17)^2 + (5-9.17)^2 + ... + (25-9.17)^2] = 54.81$

   $\sigma = 7.4$

   (b) $P(20) = \frac{1}{7.4\sqrt{2\pi}} e^{-\frac{-(20-9.17)^2}{2*54.81}} = .0185$

   (c) This is equivalent to the probability of each event happening simultaneously $= P(2) * P(5) * P(7) * P(7) * P(9) * P(25)$
   $= \frac{1}{7.4\sqrt{2\pi}} e^{-\frac{-(2-9.17)^2}{2*54.81}} + ... + \frac{1}{7.4\sqrt{2\pi}} e^{-\frac{-(25-9.17)^2}{2*54.81}}$

   $= 1.2E - 9$

(d) The probability would be greater because you're replacing two values (one of which was a relative outlier) with values closer to the mean, in a normal distribution this corresponds to a much greater probability for each discrete value and thus a higher probability all together.

(e) $\mu_y = \frac{4+4+5+6+8+10}{6} = 6.17$

$cov(X,Y) = \sum_{i=1}^{m} \frac{(x_i - \mu_x)(y_i - \mu_y)}{m} =$

(f) MSE measures the average squared value of the line-of-fit, and the point's it's trying to fit. Covariance is a value representing how two variables change with respect to one another. The MSE should be a larger value when there's is a very low covariance, meaning there will be less of a real "fit" to any prediction when one variable doesn't predict another. If there is a very high covariance however, that implies a strong relation between the variables, and the MSE should me lower as a representation of being able to predict one based on another.