Machine Learning

Linear Regression

1. Some equations

$$
\begin{aligned}
\nabla_A \mathrm{tr}(AB) = \nabla_A \mathrm{tr}(BA) = B^T \\
\nabla_{A^T} f(A) = (\nabla_A f(A))^T \\
\nabla_A \mathrm{tr}(ABA^T C) = CAB + C^T AB^T \\
\nabla_A |A| = |A| (A^{-1})^T
\end{aligned}
$$

2. Model

input:

$$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})$$

hypothesis:

$$
\begin{aligned}
h_\theta(x) &= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n \\
&= \sum_{i=0}^{n} \theta_i x_i = \theta^T x
\end{aligned}
$$

cost function:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 = \frac{1}{2} (X\theta - y)^T (X\theta - y)$$

derivative:

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta (\frac{1}{2} \theta^T X^T X \theta - y^T X \theta - \frac{1}{2} y^T y) \\
&= X^T X \theta - X^T y
\end{aligned}
$$

solution:

$$\nabla_\theta J(\theta) = 0 \Longrightarrow X^T X\theta = X^T y \Longrightarrow \theta = (X^T X)^{-1} X^T y$$

3. Locally weighted linear regression

weights:

$$\omega_i = \exp(-\frac{(x^{(i)} - x)^2}{2k^2})$$

cost function:

$$\begin{aligned} J(\theta) &= \frac{1}{2}\sum_{i=1}^{m} \omega_i (h_\theta(x^{(i)}) - y^{(i)})^2 \\ &= \frac{1}{2}(WX\theta - Wy)^T (X\theta - y) \end{aligned}$$

derivative:

$$\nabla_\theta J(\theta) = X^T W X\theta - X^T W y$$

solution:

$$\begin{aligned} \nabla_\theta J(\theta) = 0 &\Longrightarrow X^T W X\theta = X^T W y \\ &\Longrightarrow \theta = (X^T W X)^{-1} X^T W y \end{aligned}$$

4. Ridge regression

cost function:

$$\begin{aligned} J(\theta) &= \frac{1}{2}\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{1}{2}\lambda \sum_{i=0}^{n} \theta_i^2 \\ &= \frac{1}{2}(X\theta - y)^T (X\theta - y) + \frac{1}{2}\lambda \theta^T \theta \end{aligned}$$

derivative:

$$\nabla_\theta J(\theta) = (X^T X + \lambda I)\theta - X^T y$$

solution:

$$
\begin{aligned}
\nabla_\theta J(\theta) = 0 &\implies (X^T X + \lambda I)\theta = X^T y \\
&\implies \theta = (X^T X + \lambda I)^{-1} X^T y
\end{aligned}
$$