Machine Learning

Linear Regression

1. Some equations

$$egin{aligned}
abla_A ext{tr}(AB) &=
abla_A ext{tr}(BA) = B^T \
abla_{A^T} f(A) &= (
abla_A f(A))^T \
abla_A ext{tr}(ABA^TC) &= CAB + C^TAB^T \
abla_A |A| &= |A|(A^{-1})^T \end{aligned}$$

2. Model

input:

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

hypothesis:

$$egin{aligned} h_{ heta}(x) &= heta_0 + heta_1 x_1 + heta_2 x_2 + \dots + heta_n x_n \ &= \sum_{i=0}^n heta_i x_i = heta^T x \end{aligned}$$

cost function:

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

derivative:

$$egin{aligned}
abla_{ heta} J(heta) &=
abla_{ heta} (rac{1}{2} heta^T X^T X heta - y^T X heta - rac{1}{2} y^T y) \ &= X^T X heta - X^T y \end{aligned}$$

solution:

$$abla_{ heta}J(heta)=0\Longrightarrow X^TX heta=X^Ty\Longrightarrow heta=(X^TX)^{-1}X^Ty$$

3. Locally weighted linear regression

weights:

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

cost function:

$$egin{align} J(heta) &= rac{1}{2} \sum_{i=1}^m \omega_i (h_ heta(x^{(i)}) - y^{(i)})^2 \ &= rac{1}{2} (WX heta - Wy)^T (X heta - y) \end{split}$$

derivative:

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

solution:

$$abla_{ heta}J(heta) = 0 \Longrightarrow X^TWX heta = X^TWy \ \Longrightarrow heta = (X^TWX)^{-1}X^TWy$$

4. Ridge regression

cost function:

$$egin{split} J(heta) &= rac{1}{2} \sum_{i=1}^m (h_ heta(x^{(i)}) - y^{(i)})^2 + rac{1}{2} \lambda \sum_{i=0}^n heta_i^2 \ &= rac{1}{2} (X heta - y)^T (X heta - y) + rac{1}{2} \lambda heta^T heta \end{split}$$

derivative:

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

solution:

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