The problem of overfitting

Overfitting 과적합

ex1. $\theta_0 + \theta_1 x$ Underfit, High Bias 편향

ex2. $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$ Overfit, High Variance 분산

「If we have too many features,
the learned hypothesis may fit the training set very well,
but fail to generalize to new examples.」

과적합 피하려면,

1. Reduce number of features
   - Manually select which features to keep
   - Model selection algorithm
2. Regularization
   - Keep all the features, but reduce magnitude/values of parameters θ_j
   - Works well when we have a lot of features, each of which contributes a bit to predicting y.

Regularization

↳ Small values for parameters $\theta_0, \theta_1, \cdots, \theta_n$

- "Simpler" hypothesis
- Less prone to overfitting

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right]$$

regularization parameter

- controls a tradeoff btw. two different goals.

① fitting the trainset well

② keeping the parameters small

Regularized linear regression

Gradient Descent (θ_0는 penalize X)

$$\theta_j := \theta_j - \alpha \cdot \left[\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}\theta_j\right]$$

$\frac{\partial}{\partial \theta_j} J(\theta)$ <regularized>

같다

$$\theta_j := \theta_j \boxed{\left(1 - \alpha \cdot \frac{\lambda}{m}\right)} - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

less than 1, θ_j는 규제 시 더 작게 update

Regularized logistic regression

$$J(\theta) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] \underline{+ \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2}$$

↳ means to explicitly exclude the bias term, θ_0

Gradient descent

Repeat {

$$\theta_0 := \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha\underbrace{\left[\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m}\theta_j\right]}_{\frac{\partial}{\partial\theta_j}J(\theta)}$$

$(j = \cancel{0}, 1, 2, 3, \ldots, n)$

$\theta_1 \ldots \theta_n$

}

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$