2_Linear classification

Like linear regression, the whole procedure is just make a **hypothesis**(basis function class), write out the **Loss Function**, use **gradient descent** to find the solution of weights.

Since it's a classification problem now, the output should be probability, in the range of [0,1]. So, Sigmoid function works.

The following is bases on 2 class.

2.1 Hypothesis

$p(y=1|x)=h_\theta(x)=g(\theta^Tx)=g(\theta_0+\theta_1x_1+\ldots+\theta_dx_d)$, where g(z) is the sigmoid(z) function $\frac{1}{1+exp(-z)}$

2.2 Loss Function

Since we knew that $P(y|x)=[p(y=1|x)]^y[p(y=0|x)]^{1-y}$, we can get $p(D|\theta)=\Pi_{i=1}^mP(y^i|x^i;\theta)$.

Applying the same strategy in MAP, the NLL, we can get the loss function:

$$J(\theta)=-\frac{1}{m}\sum_{i=1}^m y^ilogh_\theta(x^i)+(1-y^i)log(1-h_\theta(x^i))$$

And of course, we can add an regularization term:

$$L2:\ \ J(\theta)=-\frac{1}{m}\sum_{i=1}^m y^ilogh_\theta(x^i)+(1-y^i)log(1-h_\theta(x^i))+\frac{\lambda}{2m}\sum_{j=1}^d\theta_j^2$$

$$L1:\ \ J(\theta)=-\frac{1}{m}\sum_{i=1}^m y^ilogh_\theta(x^i)+(1-y^i)log(1-h_\theta(x^i))+\frac{\lambda}{m}\sum_{j=1}^d|\theta_j|$$

2.3 Multi-class

For multiclass(K classes) problem , we have K weights, each weight, θ^k, corresponds to a single class.

And the hypothesis applies softmax() function:

$$P(y=k|x;\theta)=\frac{exp(\theta^k)x}{\sum_k exp(\theta^k)x)}$$

Thus the cost function is:

$$J(\theta)=-\frac{1}{m}\sum_{i=1}^m\sum_{k=1}^K I(y^i=k)log(\frac{exp(\theta^Tx^i)}{\sum_{j=1}^k exp(\theta^{jT}x^i)})+\frac{\lambda}{2m}\sum_{j=0}^d\sum_{k=1}^K(\theta_j^k)^2$$

2.3 Generative models

The above is named **Discriminative models**. It learns $P(y|x)$ with assumption about $P(y|x)$. No priori estimation get used.

However, **Generative models** estimate parameters of $P(x|y)$, $P(y)$ with assumption on them, and then make prediction $P(y|x)$ with those parameters and Bayes rule.

This chapter is a little bit abbreviated, because I don't think it's really different from Linear regression.

I'll record more detail about Generative models in the next chapter, including the deduction of derivative.