Chapter06 Logistic Regression

1. $f_{w,b}(x) = \sigma(\sum_i w_i x_i + b)$ $\qquad f_{w,b}(x) = \sum_i w_i x_i + b$
2. $w^*, b^* = argmaxL(w,b) = argmin - lnL(w,b)$
   a) $L(w,b) = f_{w,b}(x_1) f_{w,b}(x_2)\left(1 - f_{w,b}(x_3)\right) \dots f_{w,b}(x_n)$

$$-lnL(w,b) = -(lnf_{w,b}(x_1) + lnf_{w,b}(x_2) + ln\left(1 - f_{w,b}(x_3)\right) + \cdots + lnf_{w,b}(x_n))$$

$$= -\sum[\hat{y}^n lnf_{w,b}(x^n) + (1 - \hat{y}^n) ln(1 - f_{w,b}(x^n))]$$

   b) $C(p,q) = -\sum p(x) lnq(x)$
   c) $L(f) = \sum C(f(x^n), \hat{y}^n)$ $\qquad L(f) = \frac{1}{2}\sum(f(x^n) - \hat{y}^n)^2$
3. $w_i = w_i - \eta \sum(f(x^n) - \hat{y}^n) x_i^n$
4. Logistic Regression + Square Error (x)
   Logistic Regression + Cross Entropy (v)
5. Usually, discriminative model is better than generative model
6. Benefit of generative model
   a) With the assumption of probability distribution, less training data is needed
   b) With the assumption of probability distribution, more robust to noise
7. Multi-class Classification

   $z_i = w_i \cdot x + b \qquad softmax\text{: } P(C_i|x) = e^{z_i} / \sum e^{z_i}$
8. Limitation of Logistic Regression
   a) Can not handle the linearly indivisible question
   b) Using cascading logistic regression model
      Feature Transformation + Classification