Chenxu Wang 10457625

1. Logistic regression is very interesting part for me. It is a effective way to make a classification. Especially, for its activation function: sigmoid $f(x) = \frac{1}{1+e^{-x}}$, it easily limits the input in [0, 1].

Then, it can use Gradient descent to obtain the parameter w. GD takes advantage of computer's feature to iterate datasets for compute the result. I believe Logistic Regression is the favorite part for me.

2. when we have $E(X) = 1$,

It refeclts the average value of the random varibles.

Variance is used to measure the degree of deviation between a random varible and its expectation. It can reflect the fluctuation of sample.

when we have another Random variable Y, and its $Var[Y]$, we can compare the fluctuation of these two varibles.

3.

Using chebyshev's Inequality:

X is random varible. we have

$$Pr(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

Using Markov's Inequality:

$$X > 0, \quad P(X \geq t) \leq \frac{E(X)}{t}, \quad \text{for all } (t > 0)$$

We assume, non-expert follows $B(0 \sim 1)$ $p = 0.5$

$$\therefore E_{non}(X) = p = \frac{1}{2} \qquad V_{non}(X) = p(1-p) = \frac{1}{4}$$

Expert follows $B(0 \sim 100)$ $p = 0.99$

$$E_{exp}(X) = p \cdot 100 = 99 \qquad V_{exp}(X) = p(1-p) = 0.0099$$

Using chebyshev's Inequality:

$$P(Y_i = 1) = 0.5 \qquad P(Y_i = 0) = 0.5$$

$$X = Y_1 Y_2 \cdots Y_n$$

when $X \geq E(X) + t$

Since we need the quality is as good as expert.

$$\therefore E_{non}(X) + t = 99 \Rightarrow t = 99 - E(X) = 99 - \frac{1}{2}n$$

$$\therefore P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

We assume a sufficient condition/confidence as p

$$\therefore \frac{Var_{non}(x)}{t^2} = 1-p$$

$$\therefore 1-p = \frac{\frac{1}{4}}{(99-\frac{1}{2}n)^2}$$

$$\therefore (99-\frac{1}{2}n)^2 = \frac{1}{4(1-p)}$$

since $p < 1$

$$\therefore 99-\frac{1}{2}n = \frac{1}{2}\sqrt{\frac{1}{1-p}}$$

$$n = 198 - \sqrt{\frac{1}{1-p}}$$

$$\therefore f(p) = 198 - \sqrt{\frac{1}{1-p}} \quad (f(p) \geq 0)$$

4.

①. RP is fast when computes high dimension data. and performs better when data is independent.

②. when matrix X is not a sparse matrix or k is too small. RP has a bad efficiency.

③ It will produce many different results when we use different random matrices. But they can all guarantee the accuracy. It could use a lot of memory. However, it can produce Random matrix faster compared with other method.

5. Collaborative filtering can predict a user's behavior by other users' behavior. It can filter out all empty entries in the sparse matrix.

The formulation is: $\min_{X} \|(Z - X)_{\Omega}\|_F^2 \quad \text{s.t. } \mathrm{rank}(X) \le r$

Drawbacks:

It has a bad quality for new users. & new items.

It can be solved by new formulation:

$X = U \cdot V^T$

$$\min_{U.V} \|(Z - UV^T)_{\Omega}\|_F^2 + \lambda(\|U\|_F^2 + \|V\|_F^2)$$