Logistic Regression

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Learn the logistic regression classifier
  - Understand why the logistic regression is better suited than the linear regression for classification tasks
  - Understand the logistic function
  - Understand the logistic regression classifier
  - Understand the approximation approach for the open form solutions
- Learn the gradient descent algorithm
  - Know the tailor expansion
  - Understand the gradient descent/ascent algorithm
- Learn the different between the naïve Bayes and the logistic regression
  - Understand the similarity of the two classifiers
  - Understand the differences of the two classifiers
  - Understand the performance differences

NAÏVE BAYES VS. LOGISTIC REGRESSION

Gaussian Naïve Bayes

- We want to compare the performance of the two classifiers
  - Logistic regression handles the continuous features
  - Why not naïve Bayes?
- Naïve Bayes Classifier Function
  - $f_{NB}(x) = argmax_{Y=y} P(Y = y) \prod_{1 \le i \le d} P(X_i = x_i | Y = y)$
- What-if the feature is a continuous random variable?
  - We can assume that the variable follows the Gaussian distribution with the mean of μ and the variance of σ^2
    - $P(X_i | Y, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$
  - In addition, let's use more shortened terms
    - $P(Y = y) = \pi_1$
  - $P(Y) \prod_{1 \le i \le d} P(X_i | Y) = \pi_k \prod_{1 \le i \le d} \frac{1}{\sigma_k^i C} \exp(-\frac{1}{2}\left(\frac{X_i - \mu_k^i}{\sigma_k^i}\right)^2)$

Derivation to Logistic Regression (1)

- Derivation from the naïve Bayes to the logistic regression
  - $P(Y)\prod_{1\le i\le d} P(X_i|Y) = \pi_k \prod_{1\le i\le d} \frac{1}{\sigma_k^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_k^i}{\sigma_k^i}\right)^2)$
- With naïve Bayes assumption
  - $P(Y=y|X) = \frac{P(X|Y=y)P(Y=y)}{P(X)} = \frac{P(X|Y=y)P(Y=y)}{P(X|Y=y)P(Y=y)+P(X|Y=n)P(Y=n)}$

$$= \frac{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y)}{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y) + P(Y=n)\prod_{1\le i\le d} P(X_i|Y=n)}$$

Derivation to Logistic Regression (2)

- With naïve Bayes assumption
  - $P(Y=y|X) = \frac{P(X|Y=y)P(Y=y)}{P(X)} = \frac{P(X|Y=y)P(Y=y)}{P(X|Y=y)P(Y=y)+P(X|Y=n)P(Y=n)}$

$$= \frac{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y)}{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y) + P(Y=n)\prod_{1\le i\le d} P(X_i|Y=n)}$$

  - $P(Y=y|X) = \frac{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2) + \pi_2 \prod_{1\le i\le d} \frac{1}{\sigma_2^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}$

$$= \frac{1}{1 + \frac{\pi_2 \prod_{1\le i\le d} \frac{1}{\sigma_2^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}}$$

Derivation to Logistic Regression (3)

- Assuming the same variable of the two classes, σ_2^i=σ_1^i
  - $$P(Y=y|X) = \frac{1}{1+\frac{\pi_2 \prod_{1\le i\le d}\frac{1}{\sigma_2^i C}\exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}{\pi_1 \prod_{1\le i\le d}\frac{1}{\sigma_1^i C}\exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}} = \frac{1}{1+\frac{\pi_2 \prod_{1\le i\le d}\exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}{\pi_1 \prod_{1\le i\le d}\exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}}$$

$$= \frac{1}{1+\frac{\pi_2\exp(-\sum_{1\le i\le d}\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\})}{\pi_1\exp(-\sum_{1\le i\le d}\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\})}}$$

$$= \frac{1}{1+\frac{\exp(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right\}+log\pi_2)}{\exp(-\sum_{1\le i\le d}\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\}+log\pi_1)}}$$

Derivation to Logistic Regression (4)

- Assuming the same variable of the two classes, $\sigma_2^i = \sigma_1^i$
  - $P(Y=y|X) = \dfrac{1}{1+\exp(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right\}+log\pi_2+\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right\}-log\pi_1)}$
  - $= \dfrac{1}{1+\exp(-\frac{1}{2(\sigma_1^i)^2}\sum_{1\le i\le d}\left\{\left(X_i-\mu_1^i\right)^2-\left(X_i-\mu_2^i\right)^2\right\}+log\pi_2-log\pi_1)}$
  - $= \dfrac{1}{1+\exp(-\frac{1}{2(\sigma_1^i)^2}\sum_{1\le i\le d}\left\{2\left(\mu_2^i-\mu_1^i\right)X_i+{\mu_2^i}^2-{\mu_2^i}^2\right\}+log\pi_2-log\pi_1)}$