Chapter 5

김선
서울대학교 컴퓨터 공학부
생물정보 연구소

5

로지스틱 회귀 & ROC 곡선
(Logistic regression & ROC curve)

김선
서울대학교 컴퓨터 공학부
생물정보 연구소

Remind: Linear regression

- Linear regression can predict values in $[-\infty, \infty]$.
- What if target values are not real number?
  - Gender, Race, …
  - Cancer patient or not

Logistic regression

- Logistic regression solves this problem by defining probability of target values.
  - Whereas linear regression directly model target values.
- If there exist two target values
  - Assume Bernoulli distribution.
- If there exist multiple target values (> 2)
  - Assume Multinomial distribution.

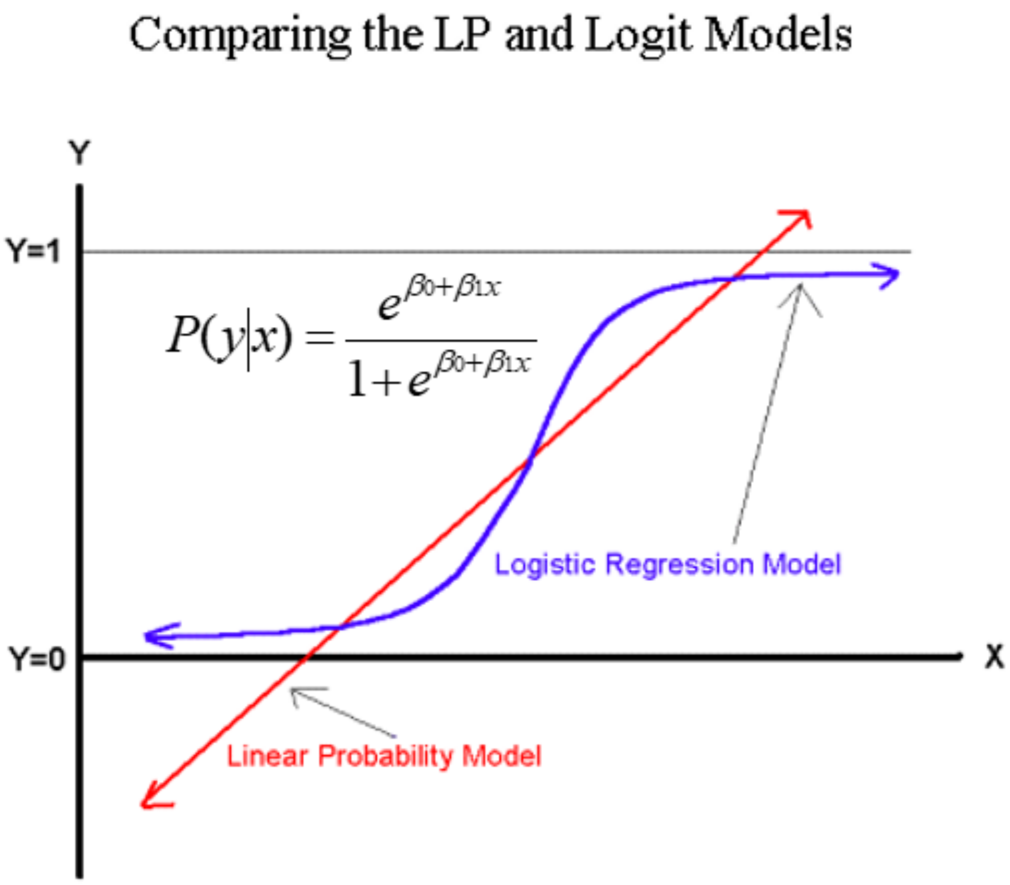
Logit

- Odds (A to B): likelihood of occurrence event A than event B.
  - $= \frac{P(A)}{P(B)}$
- In case of Bernoulli Distribution
  - $P(Y = k) = p^k(1 - p)^{1-k}$
    - $P(Y = 1) = p$
    - $P(Y = 0) = 1 - p$
  - $odds = \frac{P(Y=1)}{P(Y=0)} = \frac{p}{1-p}$
- Logit: log of odds
  - $logit = \ln(odds) = \ln(\frac{p}{1-p})$

Logistic regression

- We assume logit can be modeled as linear function.
  - $f(x) = w_0 + w_1 x$
  - $logit = \ln\left(\frac{p}{1-p}\right) = f(X)$
    - $\frac{p}{1-p} = e^{f(X)}$
    - $\left(1 + e^{f(X)}\right)p = e^{f(X)}$
    - $p = \frac{e^{f(X)}}{1+e^{f(X)}} = \frac{1}{1+e^{-f(X)}} = \frac{1}{1+e^{-w_0-w_1 x}} = P(Y = 1|X = x)$

Linear Probability vs. Logit



Logistic regression

- Problem setting
  - Data: $D = \{(x,y)^n\}_{n=1}^N$
  - Input features: $x = (x_1, .., x_k)$
  - Output: $y \in \{0, 1\}$
  - Hypothesis: $h(x;\theta) = \hat{y}$
  - Parameters: $\theta = (w_0, \dots, w_k)$

- Logistic Model
  - $\hat{y} = h(x;\theta) = \begin{cases} 1 \; if \; logit \geq 0 \\ 0 \; if \; logit < 0 \end{cases}$
  - $logit = \frac{P(y=1|x)}{P(y=0|x)}$
  - $p = P(y=1|x) = g(f(x)) = \frac{1}{1+e^{-f(x)}}$
    - $g(z)$: logistic (sigmoid) function
    - $f(x) = w_0 + \sum_{i=1}^k w_i x_i$: linear function
  - $P(y=0|x) = 1 - p$
- Likelihood
  - $P(y|x;\theta) = p^y (1-p)^{(1-y)}$

Find model parameter

- By maximum likelihood estimation.
  - $\underset{\theta}{\operatorname{argmax}}\, l(\theta|D)$
  - $l(\theta|D) = \log[L(\theta|D)] = \log\left[\prod_{n=1}^{N} p\,(y|x;\theta)\right] = \log\left[\prod_{n=1}^{N} p^{y}(1-p)^{(1-y)}\right] = \sum_{n=1}^{N}[y\log p + (1-y)\log(1-p)]$
- Numerical Solutions with approximation
  - Gradient Descent
  - Newton's Method
    - Described in the Book p.156-163
  - …

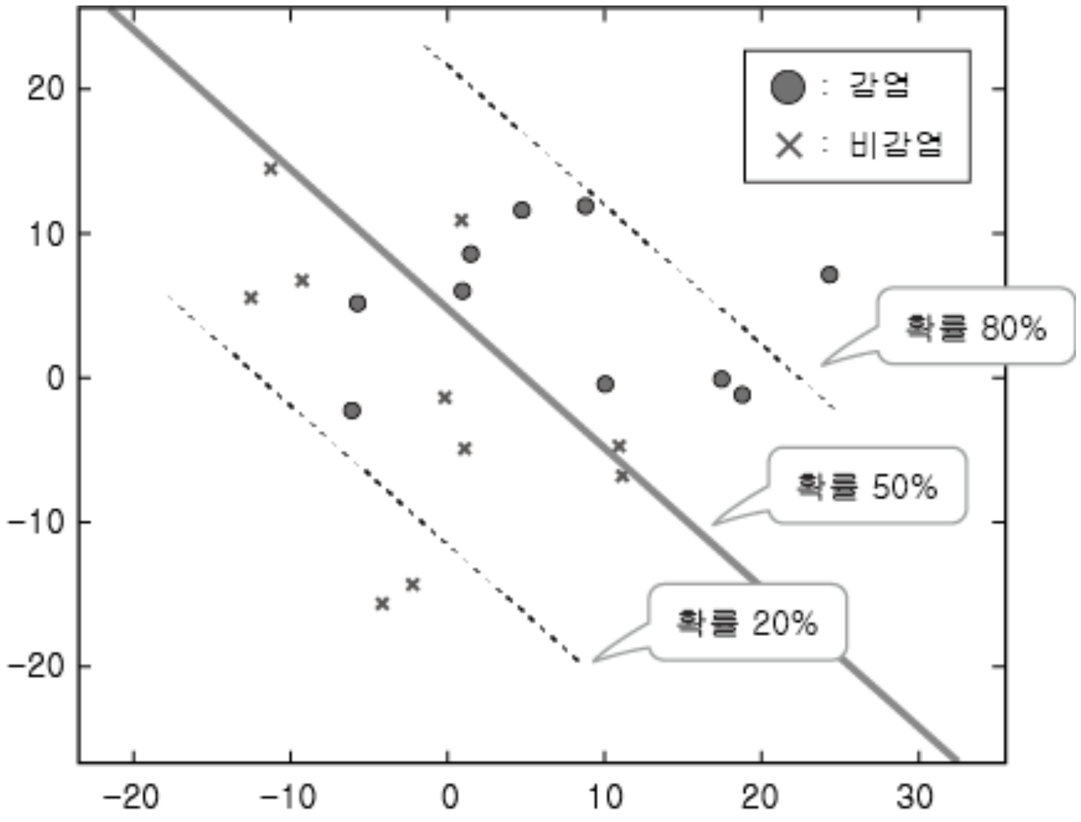
Gradient Descent

- Rather than maximize log likelihood, let's minimize negative log likelihood.
  - $\underset{\theta}{\operatorname{argmax}}\, l(\theta|D) = \underset{\theta}{arg\min} -l(\theta|D) = \underset{\theta}{arg\min} - \sum_{n=1}^{N}[y \log p + (1-y)\log(1-p)]$
- Gradient descent
  - $J(\theta) = -l(\theta|D) = -\sum_{n=1}^{N}[y \log p + (1-y)\log(1-p)]$
  - $\frac{\partial}{\partial w_i} J(\theta) = \cdots$
  - $w_i^{t+1} = w_i^t - \alpha \frac{\partial}{\partial w_i} J(\theta), \quad i = 0, \dots, k$

Model Evaluation

- From logistic model, we can calculate probability of event to be occurred for given data $\equiv P(y = 1|x)$.
- Ideally, we predict event occurred when $P(y = 1|x) > 0.5$ because $P(y = 1|x) > P(y = 0|x)$ in this case.
- For real world data, however, cutoff value should be adjusted.
  - Example) When there exists small number of positives.
    - Is 90% accuracy good when 99% of the instances are negative?
  - Example) Cancer diagnosis.
    - Is it reasonable to determine the risk of cancer when $P(y = cancer|x) > 0.5$?

Model Evaluation



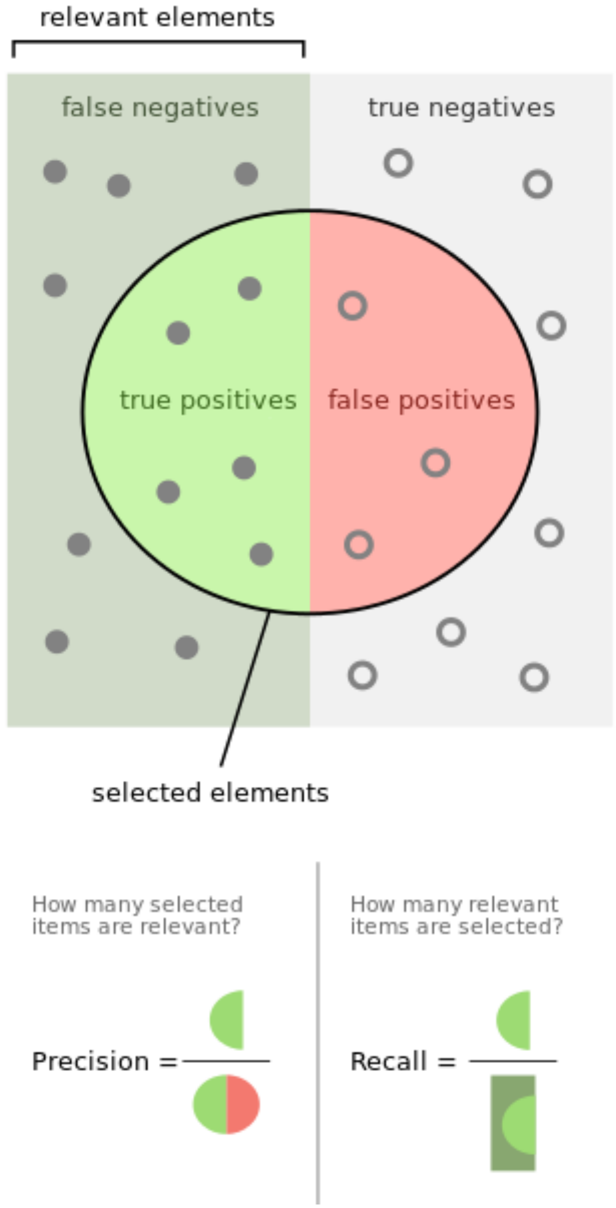
Model Evaluation

- For a threshold α, we denote
  - Positive as data which we predict as positive i.e. $P(y = 1|x) > \alpha$.
  - Negative as data which we predict as negative i.e. $P(y = 1|x) < \alpha$.
  - True as correct prediction.
  - False as wrong prediction.
- Therefore,
  - True positive (hit): Correct prediction of the positive class.
  - True negative (correct rejection): Correct prediction of the negative class.
  - False positive (false alarm): Predict originally negative class as positive.
  - False negative (miss): Predict originally positive class as negative.

Model Evaluation

		실제 정답	
		Positive	Negative
예측 결과	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

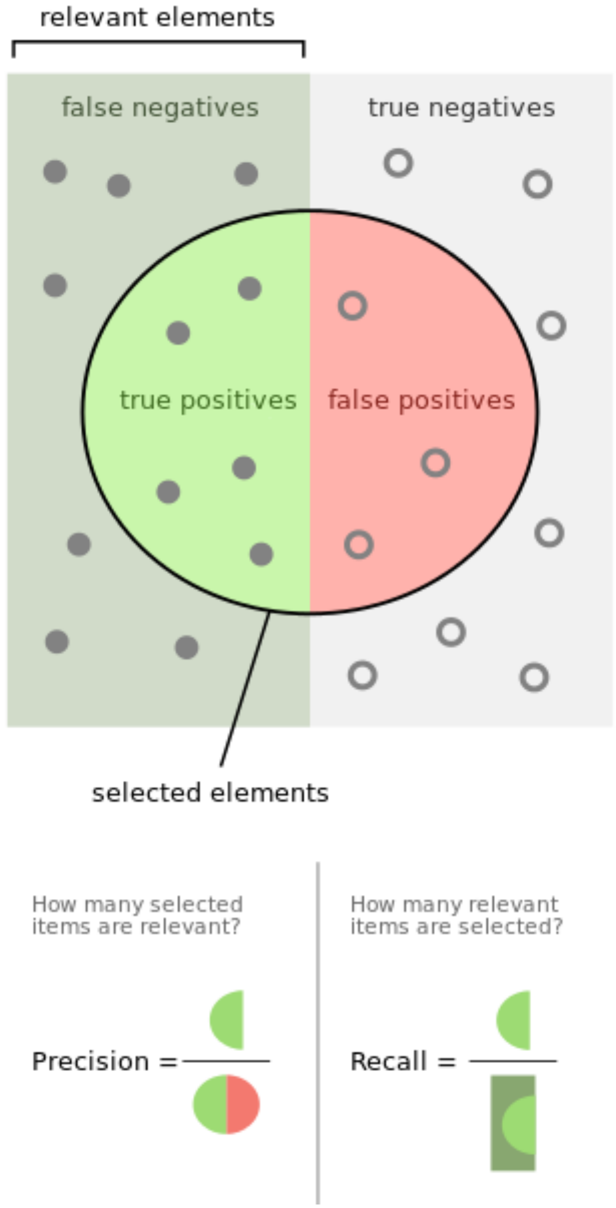
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$: ratio of correct predictions.
- $Precison = \frac{TP}{TP+FP}$: ratio of correctly predicted positives to predicted positives.
- $Recall = \frac{TP}{TP+FN}$: ratio of correctly predicted positives to original positives.



Model Evaluation

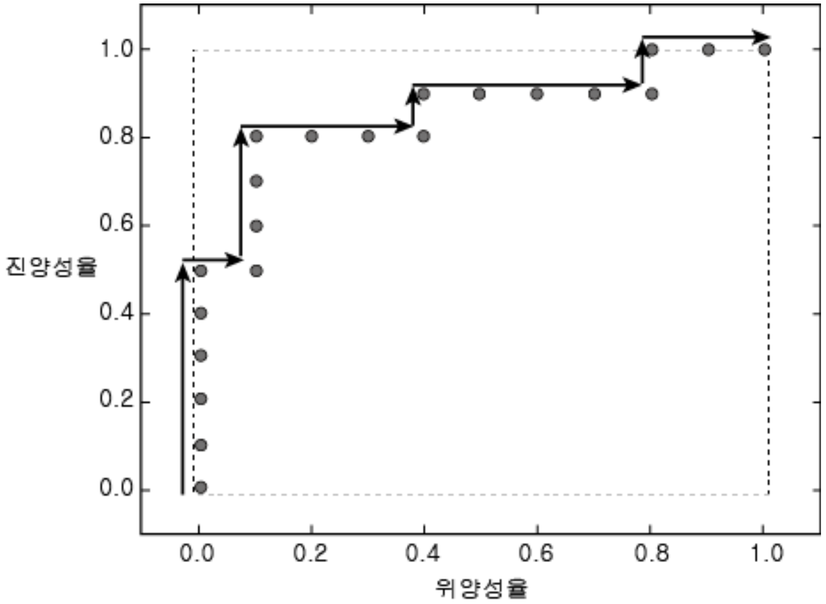
		실제 정답	
		Positive	Negative
예측 결과	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

- $Sensitivity = True\ Positive\ Rate\ (TPR) = Recall\ = \frac{TP}{TP+FN}$: ratio of correctly predicted positives to original positives.
- $Specificity = True\ negative\ rate\ (TNR) = \frac{TN}{TN+FP}$: ratio of correctly predicted negatives to original negatives.
- $False\ positive\ rate\ (FPR) = \frac{FP}{TN+FP} = 1 - TNR$: ratio of wrongly predicted positives to original negatives.



ROC curve

- Receiver operating characteristic (ROC) curve
  - plotting TPR vs. FPR as varying threshold.
  - Performance of the model for any choice of threshold.



Area Under the Curve (AUC)

- AUC: Are under the ROC curve.
  - measures the quality of classifier.
    - $AUC = 0.5$: random classifier.
    - $AUC = 1$: perfect classifier.