Statistical Machine Learning

3주차

담당: 13기 박주영

Classification

Generalized Linear Model

1. **Random component** identifies the response variable Y and its probability distribution;
2. **Linear predictor** specifies explanatory variables used in a linear predictor function; and
3. **Link function** specifies the function of $E(Y)$ that the model equates to the systematic component.

Logistic Regression

$$Y_i \overset{ind}{\sim} \text{Bernoulli}(\pi_i(\mathbf{X}_i)) \quad \text{where} \quad E[Y_i] = \pi_i(\mathbf{X}_i)$$

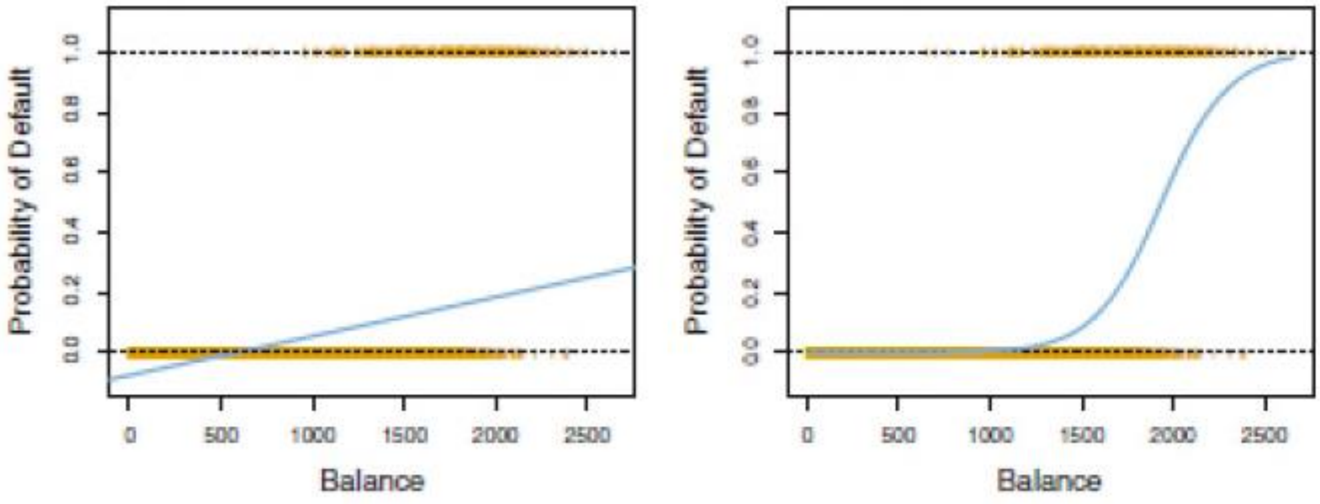
$$\log\left(\frac{\pi_i(\mathbf{X}_i)}{1-\pi_i(\mathbf{X}_i)}\right) = \beta_0 + \beta_1 X_{1\mathrm{i}} + \beta_2 X_{2\mathrm{i}} + \cdots + \beta_p X_{p\mathrm{i}}$$

Logistic Regression

$$\mathrm{P}(Y_i = 1|\mathbf{X}_i) = \pi_i(\mathbf{X}_i) = \frac{e^{\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots+\beta_p X_{pi}}}{1+e^{\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\cdots+\beta_p X_{pi}}}$$

$$= \frac{e^{\boldsymbol{\beta}^T \mathbf{x}_i}}{1+e^{\boldsymbol{\beta}^T \mathbf{x}_i}} = \frac{1}{1+e^{-\boldsymbol{\beta}^T \mathbf{x}_i}} \quad \text{(sigmoid function)}$$

Logistic Regression



Logistic Regression

- How to Estimate? $\underset{\boldsymbol{\beta}}{argmax}\ L(\boldsymbol{\beta})$

$$L(\boldsymbol{\pi};\mathbf{X}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$l(\boldsymbol{\pi};\mathbf{X}) = \sum_{i=1}^{n} [y_i \log \pi_i + (1-y_i)\log(1-\pi_i)]$$

Softmax function

- 이항 반응변수: Logistic Regression Model – Sigmoid function
- 다항 반응변수
    - 명목형: 일반화 로짓 모형 – Softmax function
    - 순서형: 누적 로짓 모형

Loss function for Classification

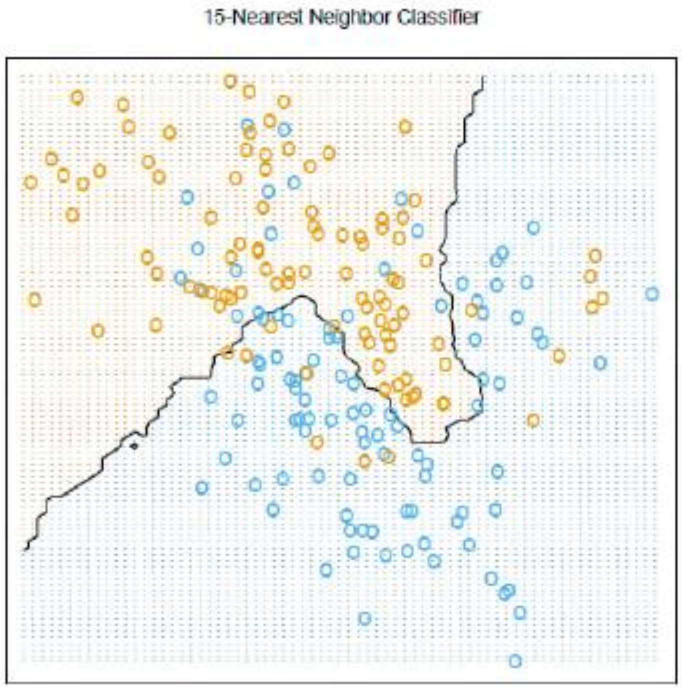
- Categorical Cross Entropy

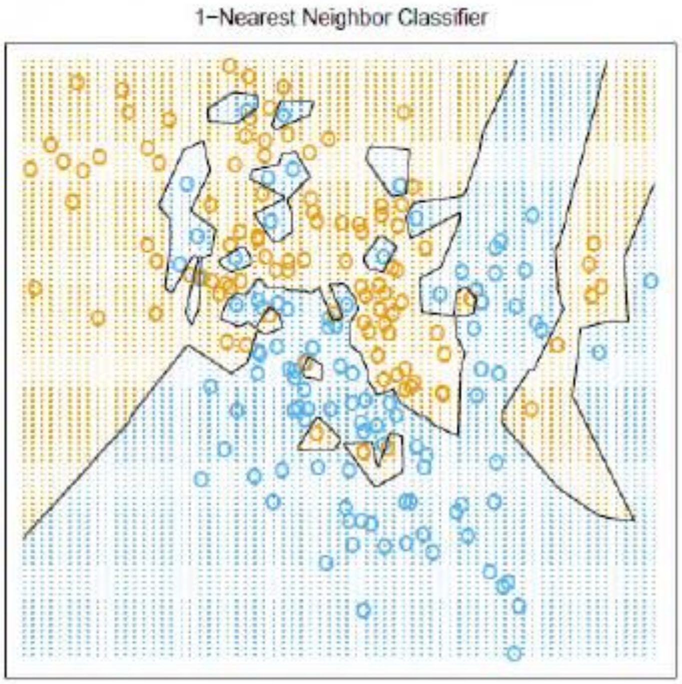
$$CE_i = -\sum_{k=1}^{C} y_{ik} \log \pi_i(k)$$

- Binary Cross Entropy

$$\begin{aligned} CE_i &= -[y_{i1} \log \pi_i(1) + y_{i0} \log \pi_i(0)] \\ &= -[y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i\ )] \end{aligned}$$

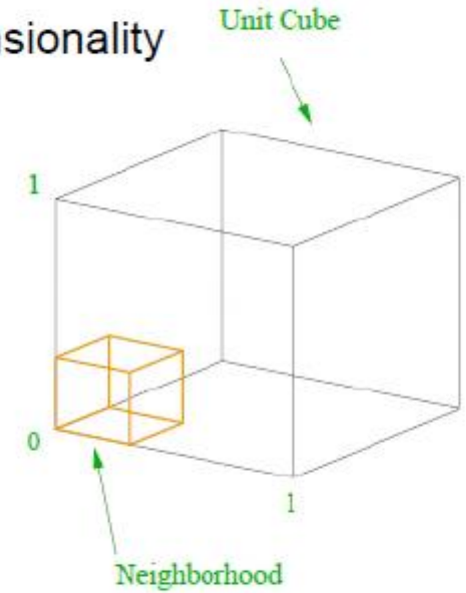
KNN Classifier

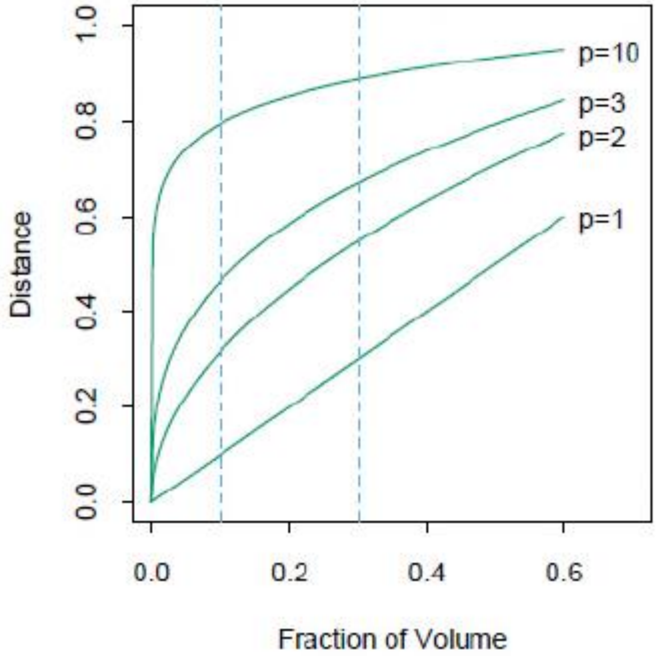




KNN Classifier

- Curse of dimensionality





KNN Classifier

- Distance measure

$$d(\mathbf{u},\mathbf{v}) = \left(\sum |u_i - v_i|^2\right)^{\frac{1}{2}} = ||\mathbf{u}-\mathbf{v}||_2$$ *Euclidean (L2 norm)*

$$d(\mathbf{u},\mathbf{v}) = \sum |u_i - v_i| = ||\mathbf{u}-\mathbf{v}||_1$$ *Manhattan (L1 norm)*

$$d(\mathbf{u},\mathbf{v}) = \left(\sum |u_i - v_i|^p\right)^{\frac{1}{p}} = ||\mathbf{u}-\mathbf{v}||_p$$ *Minkowski (Lp norm)*

$$d(\mathbf{u},\mathbf{v}) = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$ *Mahalanobis Distance*

Imbalanced data

SMOTE	ADASYN
Oversampling	Oversampling
K- nearest neighbors	SMOTE의 개선된 버전
$x_{syn} = x_i + \lambda\ (x_k - x_i),\ x_k \in S_i$	표본 수에 대한 weight를 통해 추출

Kernel Density Estimator

- Kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j))$$

Gaussian Kernel (Radial Basis function)

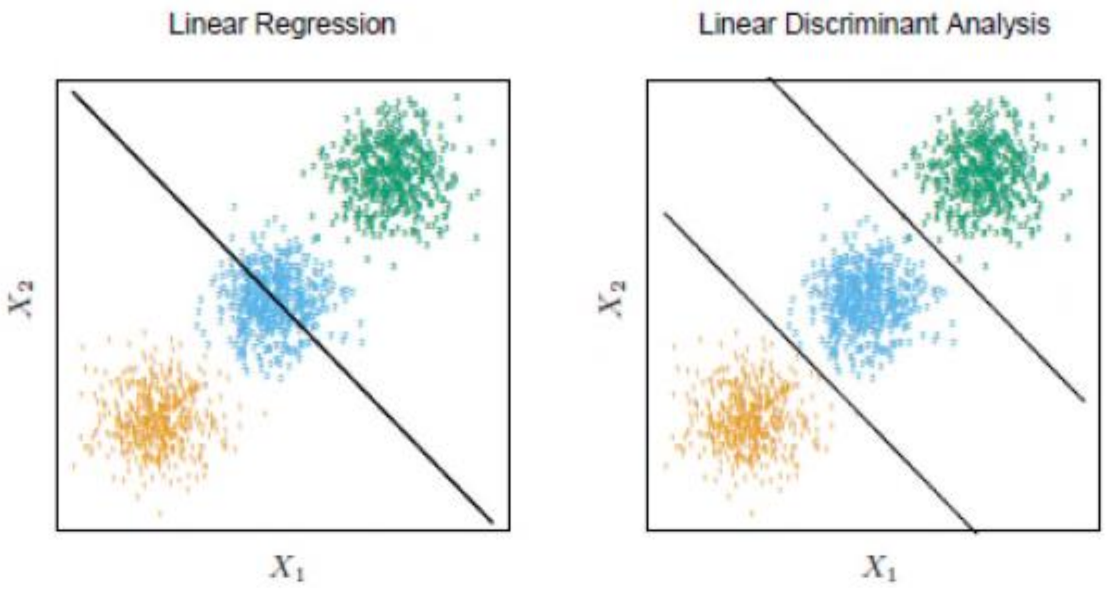
$$K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i^T \mathbf{x}_j)^p$$

polynomial Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(k_1 \mathbf{x}_i^T \mathbf{x}_j + k_2)$$

Sigmoid Kernel

Discriminant Analysis



Naïve Bayes Classifier

$$P(Y_i = k|\mathbf{X}_i) = \frac{P(\mathbf{X}_i|k)P(k)}{\sum_k P(\mathbf{X}_i|k)P(k)}$$

Bayes' Theorem

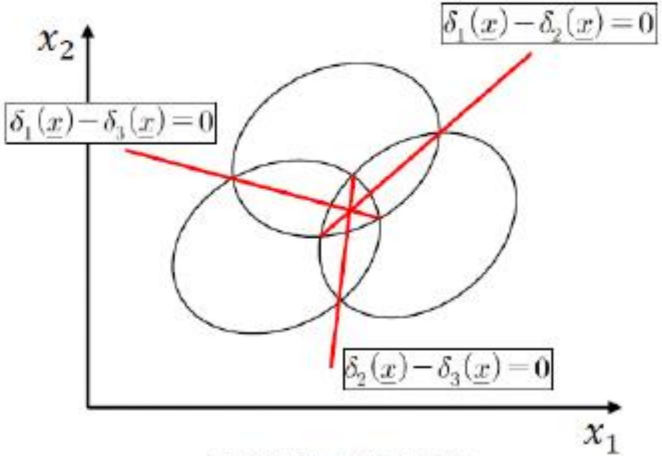
$$\text{where} \quad P(\mathbf{X}_i|k) = \prod_j^p P(X_{ij}|k)$$

Linear Discriminant Analysis

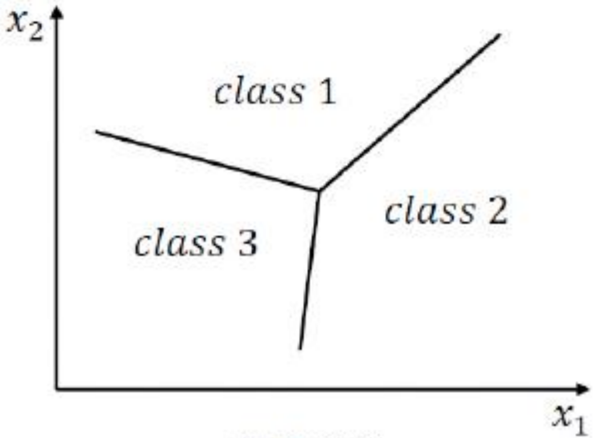
$$P(Y_i = k|\mathbf{X}_i) = \frac{P(\mathbf{X}_i|k)P(k)}{\sum_k P\,(\mathbf{X}_i|k)P(k)} \qquad Bayes'\ Theorem$$

$$\text{where} \quad P(\mathbf{X}_i|k) \sim N_p(\boldsymbol{\mu}_k, \Sigma)$$

Linear Discriminant Analysis



<판별 함수의 경계>



<판별 영역>

Linear Discriminant Analysis

$$\text{IF } P(Y_i = k|\mathbf{X}_i) > P(Y_i = l|\mathbf{X}_i) \rightarrow \textit{estimate class of } Y_i \textit{ to } k$$

$$\log\frac{P(Y_i = k|\mathbf{X}_i)}{P(Y_i = l|\mathbf{X}_i)} = \delta_k(\mathbf{X}_i) - \delta_l(\mathbf{X}_i)$$

$$\textit{where} \quad \delta_k(\mathbf{X}_i) = \mathbf{X}_i^T \Sigma^{-1} \boldsymbol{\mu}_k - \frac{1}{2}\boldsymbol{\mu}_k^T \Sigma^{-1} \boldsymbol{\mu}_k + \log P(k)$$

Quadratic Discriminant Analysis

$$P(Y_i = k|\mathbf{X}_i) = \frac{P(\mathbf{X}_i|k)P(k)}{\sum_k P\,(\mathbf{X}_i|k)P(k)} \qquad \textit{Bayes' Theorem}$$

$$\text{where} \quad P(\mathbf{X}_i|k) \sim N_p(\boldsymbol{\mu}_k, \Sigma_k)$$

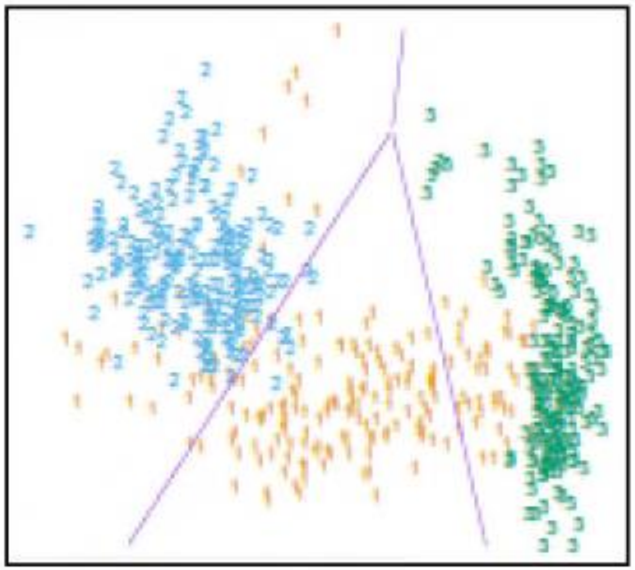
Quadratic Discriminant Analysis

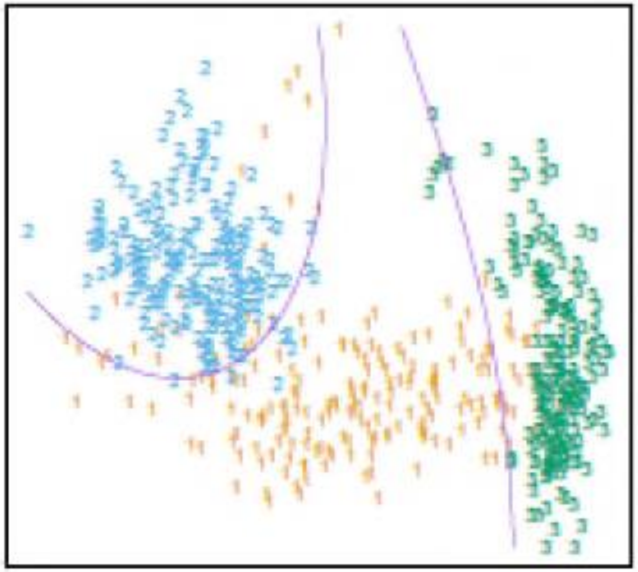
$$\text{IF } P(Y_i = k|\mathbf{X}_i) > P(Y_i = l|\mathbf{X}_i) \ \rightarrow \ \textit{estimate class of } Y_i \textit{ to } k$$

$$\log\frac{P(Y_i = k|\mathbf{X}_i)}{P(Y_i = l|\mathbf{X}_i)} = \delta_k(\mathbf{X}_i) - \delta_l(\mathbf{X}_i)$$

$$\textit{where } \ \delta_k(\mathbf{X}_i) = -\frac{1}{2}\log|\Sigma_k| \ -\frac{1}{2}(\mathbf{X}_i - \boldsymbol{\mu}_k)^T \Sigma_k{}^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_k) + \log P(k)$$

LDA and QDA





What is Regression?