Lecture 05:

Classification:

Naïve Bayes, &

Discriminant Analysis

Rawls Profess of MIS

Jaeki Song, Ph.D.

Naïve Bayes Classifier

- Applied to data with categorical predictors
- Thomas Bayes 91702-1761)
  - Approach
    - Find all of the training record with the same predictor profile (e.g., records having the same predictor values)
    - Determine what classes the records belong to
    - Assign that class to the new record
  - Question
    - What is the propensity of belonging to the class of interst?

Bayes Classifier

- A probabilistic framework for solving classification problems
- Conditional Probability: $$P(C \mid A) = \frac{P(A, C)}{P(A)}$$ $$P(A \mid C) = \frac{P(A, C)}{P(C)}$$
- Bayes theorem: $$P(C \mid A) = \frac{P(A \mid C) P(C)}{P(A)}$$

Bayesian Classifiers

- Y (conditional probability)

  $$f(y|x) \propto f(y) f(x|y), y = 1, \ldots\ldots, J$$

  , where f(y|x) → posterior distribution of x, f(y) → prior distribution of Y.

  - conditional independence

    $$P(X_1 = x_1 | X_2, Y) = P(X_1 = x_1 | Y)$$

Naïve Bayes Classifier

- Formula:
  - Calculating the probability that a record with a given set of predictor values $(x_1, \dots \dots \dots \dots, x_p)$ belongs to class C_1 among m classes

$$P\ (C_1|x_1, \dots \dots \dots \dots, x_p) = \frac{P(C_1)[P(x_1|C_1)P(x_2|C_1) \cdots\cdots P(x_p|C_1)]}{P(C_1)[P(x_1|C_1)P(x_2|C_1) \cdots\cdots P(x_p|C_1)] + \cdots\cdots + P(C_m)[P(x_1|C_m)P(x_2|C_m) \cdots\cdots P(x_p|C_m)]}$$

Naïve Bayes Classifier

Company	Prior Legal Issue	Company Size	Status
1	Y (Yes)	S (Small)	T (Truthful)
2	N (No)	S	T
3	N	L (Large)	T
4	N	L	T
5	N	S	T
6	N	S	T
7	Y	S	F (Fraudulent)
8	Y	L	F
9	N	L	F
10	Y	L	F

Naïve Bayes Classifier

$$\mathrm{P}(\mathit{fraudulent}|\mathrm{PriorLegal} = \mathrm{y}, \mathrm{Size} = \mathrm{S}) = \frac{\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{4}{10}\right)}{\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{4}{10}\right) + \left(\frac{1}{6}\right)\left(\frac{4}{6}\right)\left(\frac{6}{10}\right)} = 0.53$$

$$\mathrm{P}(\mathit{fraudulent}|\mathrm{PriorLegal} = \mathrm{y}, \mathrm{Size} = \mathrm{L}) = 0.87$$

$$\mathrm{P}(\mathit{fraudulent}|\mathrm{PriorLegal} = \mathrm{n}, \mathrm{Size} = \mathrm{S}) = 0.07$$

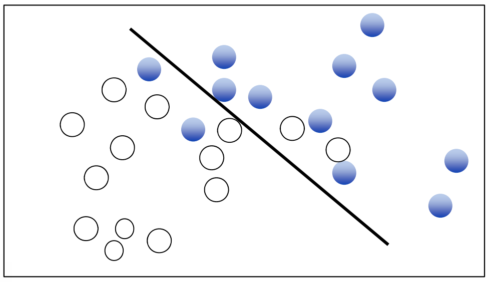
$$\mathrm{P}(\mathit{fraudulent}|\mathrm{PriorLegal} = \mathrm{n}, \mathrm{Size} = \mathrm{L}) = 0.31$$

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Independence assumption may not hold for some attributes
  - Use other techniques such as Bayesian Belief Networks (BBN)

Discriminant Analysis

- Karl Pearson
  - 1920s
  - model based approach to classification
    - discriminant function and discriminant rule



Discriminant Analysis

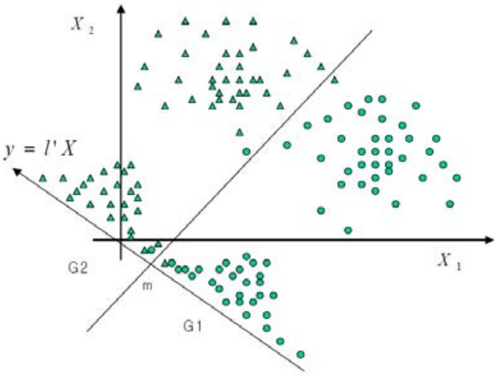
- Statistical Distance
  - Two groups
    - G1: $X_{1i} = (X_{1i1}, X_{1i2}, \ldots, X_{1ip})', \; i = 1,2,\ldots,n_1$
    - G2: $X_{2i} = (X_{2i1}, X_{2i2}, \ldots, X_{2ip})', \; i = 1,2,\ldots,n_2$

$$E[x] = \begin{cases} \mu_1 \\ \mu_2 \end{cases}$$

$$Var[x] = \Sigma$$

Discriminant Analysis

- Fisher's linear discrimination function



Discriminant Analysis

- Linear discriminant function ($Y = l'X$)
  - Fisher's linear discriminant function

$$y = l'X = (\overline{X}_1 - \overline{X}_2)' S_{pl}^{-1} X$$

,where

$$\overline{\boldsymbol{X}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} \boldsymbol{X}_{1i}, \quad \overline{\boldsymbol{X}}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} \boldsymbol{X}_{2i},$$

$$\boldsymbol{S}_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (\boldsymbol{X}_{1i} - \overline{\boldsymbol{X}}_1)(\boldsymbol{X}_{1i} - \overline{\boldsymbol{X}}_1)',$$

$$\boldsymbol{S}_2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} (\boldsymbol{X}_{2i} - \overline{\boldsymbol{X}}_2)(\boldsymbol{X}_{2i} - \overline{\boldsymbol{X}}_2)',$$

$$\boldsymbol{S}_{pl} = \left[\frac{(n_1 - 1)}{(n_1 - 1) + (n_2 - 1)}\right]\boldsymbol{S}_1 + \left[\frac{(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}\right]\boldsymbol{S}_2 = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}$$

Discriminant Analysis

- Linear discriminant function ($Y = l'X$)
  - Fisher's linear discriminant function

$$y = l'X = (\overline{X}_1 - \overline{X}_2)' S_{pl}^{-1} X$$

- discrimination criteria

$$\hat{m} = \frac{1}{2}(\overline{X}_1 - \overline{X}_2)' S_{pl}^{-1} (\overline{X}_1 + \overline{X}_2)$$

- discrimination rule

$$y > \hat{m} \rightarrow G_1$$

$$y \leq \hat{m} \rightarrow G_2$$

Discriminant Analysis

- Example

$$G_1 : \begin{pmatrix} 1 \\ 5 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 3 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$G_2 : \begin{pmatrix} -3 \\ 5 \end{pmatrix} \begin{pmatrix} -2 \\ 7 \end{pmatrix} \begin{pmatrix} -1 \\ 6 \end{pmatrix}$$

$$\overline{\boldsymbol{X}}_1 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \overline{\boldsymbol{X}}_2 = \begin{pmatrix} -2 \\ 6 \end{pmatrix}$$

$$\boldsymbol{S}_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} = \boldsymbol{S}_2$$

$$\boldsymbol{S}_{pl} \equiv \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Discriminant Analysis

$$y = \hat{l}'\boldsymbol{X} = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'\boldsymbol{S}_{pl}^{-1}\boldsymbol{X}$$

$$= (3\,,\ -2)\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = (3\,,\ -2)\begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$= (5.33\,,\ -4.67)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = 5.33X_1 - 4.67X_2$$

$$\overline{y}_1 = \hat{l}'\overline{\boldsymbol{X}}_1 = (5.33\,,\ -4.67)\begin{pmatrix} 1 \\ 4 \end{pmatrix} = -13.35$$

$$\overline{y}_2 = \hat{l}'\overline{\boldsymbol{X}}_2 = (5.33\,,\ -4.67)\begin{pmatrix} -2 \\ 6 \end{pmatrix} = -38.68$$

$$\hat{m} = \frac{1}{2}(\overline{y}_1 + \overline{y}_2) = \frac{1}{2}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'\boldsymbol{S}_{pl}^{-1}(\overline{\boldsymbol{X}}_1 + \overline{\boldsymbol{X}}_2)$$

$$= \frac{1}{2}(-13.35 - 38.68) = -26.015$$

Discriminant Analysis

City	Popu	AQI
1	11	48
1	8	20
1	12	25
1	13	32
1	6	42
1	19	25
1	21	43
1	30	24