Nearest-neighbor classifier

It's a non-parametric classifier approach.

NN classifier uses observations in L closest to the given $\underset{\sim}{x}$, closest in the feature vector space.

Let $N_K(\underline{x})$ denote the neighborhood of $\underset{\sim}{x}$ defined by the K closest points $\underset{\sim}{x}_i$s in the training set.

Obtain a "majority vote rule" for the K nearest neighbors, i.e. pts within $N_K(\underline{x})$; voting w.r.t. their class membership. The class which is the winner (having maximum counts for ~~~~ cases inside $N_K(\underline{x})$) is the assigned class.

Logistic discrimination model

Logistic regression based classification model
As opposed to the previous methods (TPM, ECM), this approach does not require strong parametric assumption regarding class conditional densities ($f(\underset{\sim}{x} | \pi_k)$).

Two-class problem

Let us first consider a 2-class classification problem
Suppose the 2 classes are π_1 & π_2

Logistic discrimination model assumes that the class conditional densities of the 2 classes satisfy

$$\log\left(\frac{f(\underset{\sim}{x} | \pi_1)}{f(\underset{\sim}{x} | \pi_2)}\right) = \beta_0 + \underset{\sim}{\beta}' \underset{\sim}{x}$$

Where; $\underset{\sim}{x}_{p\times 1} \in \mathcal{X}$ is the feature vector
$f(\underset{\sim}{x} | \pi_i)$s are class conditional densities
$\beta_0, \underset{\sim}{\beta}_{p\times 1}$: unknown deterministic constants

Note that

$$\log\left(\frac{f(\underset{\sim}{x}|\pi_1)}{f(\underset{\sim}{x}|\pi_2)}\right)$$ is referred to as the log-odds ratio.

Now

$$\log\left(\frac{f(\underset{\sim}{x}|\pi_1)}{f(\underset{\sim}{x}|\pi_2)}\right) = \beta_0 + \underset{-}{\beta}'\underset{\sim}{x}$$

$$\Leftrightarrow \log\left(\frac{\dfrac{p(\pi_1|\underset{\sim}{x})\,f(\underset{\sim}{x})}{p(\pi_1)}}{\dfrac{p(\pi_2|\underset{\sim}{x})\,f(\underset{\sim}{x})}{p(\pi_2)}}\right) = \beta_0 + \underset{-}{\beta}'\underset{\sim}{x}$$

$$\Leftrightarrow \log\left(\frac{p(\pi_2)\;p(\pi_1|\underset{\sim}{x})}{p(\pi_1)\;p(\pi_2|\underset{\sim}{x})}\right) = \beta_0 + \underset{-}{\beta}'\underset{\sim}{x}$$

$$\Leftrightarrow \frac{p(\pi_2)}{p(\pi_1)}\cdot\frac{p(\pi_1|\underset{\sim}{x})}{p(\pi_2|\underset{\sim}{x})} = e^{\beta_0 + \underset{-}{\beta}'\underset{\sim}{x}}$$

$$\Leftrightarrow \frac{p(\pi_1|\underset{\sim}{x})}{p(\pi_2|\underset{\sim}{x})} = \frac{p(\pi_1)}{p(\pi_2)}\,e^{\beta_0 + \underset{-}{\beta}'\underset{\sim}{x}}$$

$$\Leftrightarrow \quad \frac{1 - p(\pi_2 | \underset{\sim}{x})}{p(\pi_2 | \underset{\sim}{x})} = e^{\log \frac{p(\pi_1)}{p(\pi_2)} + \beta_0 + \underline{\beta}' \underset{\sim}{x}}$$

$$\left(\because p(\pi_1 | \underline{x}) = 1 - p(\pi_2 | \underline{x}) \right)$$

$$\Leftrightarrow \quad \frac{1 - p(\pi_2 | \underline{x})}{p(\pi_2 | \underset{\sim}{x})} = e^{\beta_0^* + \underline{\beta}' \underset{\sim}{x}}$$

$$\left(\beta_0^* = \beta_0 + \log \frac{p(\pi_1)}{p(\pi_2)} \right)$$

$$\Rightarrow \quad p(\pi_2 | \underline{x}) = \frac{1}{1 + e^{\beta_0^* + \underline{\beta}' \underset{\sim}{x}}}$$

and accordingly

$$p(\pi_1 | \underset{\sim}{x}) = \frac{e^{\beta_0^* + \underline{\beta}' \underset{\sim}{x}}}{1 + e^{\beta_0^* + \underline{\beta}' \underset{\sim}{x}}}.$$

Discrimination bet^n the 2-classes depend on the odds-ratio $p(\pi_1 | \underset{\sim}{x}) / p(\pi_2 | \underline{x})$.

The class assignment rule is

Assign $\underset{\sim}{x}$ to π_1 if $\dfrac{p(\pi_1 | \underline{x})}{p(\pi_2 | \underline{x})} \geq 1$

& assign $\underset{\sim}{x}$ to π_2 if $\dfrac{p(\pi_1 | \underline{x})}{p(\pi_2 | \underline{x})} < 1$

i.e. the assignment rule is

Assign $\underset{\sim}{x}$ to Π_1 if $\beta_0^* + \underset{\sim}{\beta}'\underline{x} \geq 0$

& assign $\underset{\sim}{x}$ to Π_2 if $\beta_0^* + \underline{\beta}'\underset{\sim}{x} < 0$

<u>Note</u>: The logistic discrimination is thus based on the value of $\beta_0^* + \underline{\beta}'\underset{\sim}{x}$.

<u>Remark</u>: Note that when we talk about classification in the above framework based on the value of "$\beta_0^* + \underline{\beta}'\underline{x}$", $\underset{\sim}{x}$ is the feature vector observed and we need to estimate/compute β_0^* & $\underset{\sim}{\beta}_{p\times 1}$ based on learning set. We would address this point after we look at the multiclass framework of logistic discrimination.

Logistic discrimination : multiclass

Suppose that we have c classes $\Pi_1, \ldots, \Pi_c$

Assume that log of likelihood ratios for any pair (s,c) $s=1(1)c-1$ of likelihoods is linear of the form

$$\log\left(\frac{f(\underset{\sim}{x}\mid\Pi_s)}{f(\underset{\sim}{x}\mid\Pi_c)}\right) = \beta_{s0} + \underset{\sim}{\beta}_s'\underset{\sim}{x}$$

$$s = 1(1)c-1$$

i.e the $c-1$ log-odds specifies the model for discrimination

Note that

$$\log\left(\frac{f(\underset{\sim}{x}\mid\Pi_s)}{f(\underset{\sim}{x}\mid\Pi_c)}\right) = \beta_{s0} + \underset{\sim}{\beta}_s'\underset{\sim}{x} \quad ; \quad s = 1(1)c-1$$

$$\Leftrightarrow \log\left(\frac{p(\Pi_c)}{p(\Pi_s)}\cdot\frac{p(\Pi_s\mid\underset{\sim}{x})}{p(\Pi_c\mid\underset{\sim}{x})}\right) = \beta_{s0} + \underset{\sim}{\beta}_s'\underset{\sim}{x} \quad ; \quad s = 1(1)c-1$$

$$\Leftrightarrow \frac{p(\Pi_c)}{p(\Pi_s)}\cdot\frac{p(\Pi_s\mid\underset{\sim}{x})}{p(\Pi_c\mid\underset{\sim}{x})} = e^{\beta_{s0} + \underset{\sim}{\beta}_s'\underset{\sim}{x}} \quad ; \quad s = 1(1)c-1$$

$$\Leftrightarrow \frac{p(\Pi_s\mid\underset{\sim}{x})}{p(\Pi_c\mid\underset{\sim}{x})} = e^{\beta_{s0}^* + \underset{\sim}{\beta}_s'\underset{\sim}{x}} \quad ; \quad s = 1(1)c-1$$

$$\left(\beta_{s0}^* = \beta_{s0} + \log\left(\frac{p(\Pi_s)}{p(\Pi_c)}\right)\right)$$

$$\Leftrightarrow \frac{1}{p(\pi_c|\underset{\sim}{x})} \sum_{s=1}^{c-1} p(\pi_s|\underset{\sim}{x}) = \sum_{s=1}^{c-1} e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}$$

$$\Leftrightarrow \frac{1 - p(\pi_c|\underset{\sim}{x})}{p(\pi_c|\underset{\sim}{x})} = \sum_{s=1}^{c-1} e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}$$

$$\left(\sum_{s=1}^{c} p(\pi_s|\underset{\sim}{x}) = 1 \Rightarrow \sum_{s=1}^{c-1} p(\pi_s|\underset{\sim}{x}) = 1 - p(\pi_c|\underset{\sim}{x}) \right)$$

$$\Rightarrow p(\pi_c|\underset{\sim}{x}) = \frac{1}{1 + \sum_{s=1}^{c-1} e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}}$$

$$\& \quad p(\pi_s|\underset{\sim}{x}) = p(\pi_c|\underset{\sim}{x}) \, e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}$$

i.e. $$p(\pi_s|\underset{\sim}{x}) = \frac{e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}}{1 + \sum_{s=1}^{c-1} e^{\beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x}}}.$$

$s = 1(1)c-1$

The multiclass logistic discrimination rule based on the above set of posterior of the classes is given by:

Assign $\underset{\sim}{x}$ to π_j if

$$p(\pi_j|\underset{\sim}{x}) = \max_i p(\pi_i|\underset{\sim}{x})$$

i.e Assign $\underset{\sim}{x}$ to Π_j if $(j = 1(1)c-1)$

$$\begin{cases} \beta_{j0}^{*} + \underset{\sim}{\beta}_j' \underset{\sim}{x} = \max\limits_{(i=1(1)c-1)} \{\beta_{i0}^{*} + \underset{\sim}{\beta}_i' \underset{\sim}{x}\} & \text{— (i)} \\ \text{and } \beta_{j0}^{*} + \underset{\sim}{\beta}_j' \underset{\sim}{x} > 0 & \text{— (ii)} \\ \text{otherwise assign to } \Pi_c & \end{cases}$$

[<u>Note that</u> (i) comes from the fact that

$$p(\Pi_s | \underset{\sim}{x}) = \frac{e^{\beta_{s0}^{*} + \underset{\sim}{\beta}_s' \underset{\sim}{x}}}{1 + \sum_{s=1}^{c-1} e^{\beta_{s0}^{*} + \underset{\sim}{\beta}_s' \underset{\sim}{x}}} \quad \forall\, s = 1(1)c-1$$

so $p(\Pi_j | \underset{\sim}{x}) = \max\limits_{i=1(1)c-1} p(\Pi_i | \underset{\sim}{x})$

$\Leftrightarrow$ condition (i)

Further (ii) comes from the fact that

$$p(\Pi_j | \underset{\sim}{x}) > p(\Pi_c | \underset{\sim}{x})$$

$j = 1(1)(c-1)$

$$\Leftrightarrow \frac{e^{\beta_{j0}^{*} + \underset{\sim}{\beta}_j' \underset{\sim}{x}}}{1 + \sum_{s=1}^{c-1} e^{\beta_{s0}^{*} + \underset{\sim}{\beta}_s' \underset{\sim}{x}}} > \frac{1}{1 + \sum_{s=1}^{c-1} e^{\beta_{s0}^{*} + \underset{\sim}{\beta}_s' \underset{\sim}{x}}}$$

$$\Leftrightarrow e^{\beta_{j0}^{*} + \underset{\sim}{\beta}_j' \underset{\sim}{x}} > 1$$

$\Leftrightarrow$ $\underline{\beta_{j0}^{*} + \underset{\sim}{\beta}_j' \underset{\sim}{x} > 0}$ if this does not happen then assign $\underset{\sim}{x}$ to Π_c]

In other words assignment rule is:

assign $\underset{\sim}{x}$ to π_j if $(j = 1(1)c-1)$

$$\begin{cases} p(\pi_j|\underset{\sim}{x}) > p(\pi_k|\underset{\sim}{x}) & k = 1(1)c-1 \\ & k \neq j \\ \& \ p(\pi_j|\underset{\sim}{x}) > p(\pi_c|\underset{\sim}{x}) & \text{otherwise assign } \underset{\sim}{x} \text{ to } \pi_c \end{cases}$$

$\Leftrightarrow$ assign $\underset{\sim}{x}$ to π_j $(j = 1(1)c-1)$ if

$$\beta_{j0}^* + \underset{\sim}{\beta}_j' \underset{\sim}{x} = \max_{s=1(1)c-1} \{ \beta_{s0}^* + \underset{\sim}{\beta}_s' \underset{\sim}{x} \} \ \&$$

$$\beta_{j0}^* + \underset{\sim}{\beta}_j' \underset{\sim}{x} > 0$$ otherwise assign $\underset{\sim}{x}$ to π_c

—

Parameter estimation : 2 class problem

Consider a binary variable Y with values 1 and 0. "1" associated with Π_1 population and "0" associated with Π_2 population.

Learning sample data : $(\underset{\sim}{x}_1, y_1), (\underset{\sim}{x}_2, y_2), \ldots, (\underset{\sim}{x}_n, y_n)$

$$p(\Pi_1 | \underset{\sim}{x}) = \frac{e^{\beta_0^* + \underset{\sim}{\beta}' \underset{\sim}{x}}}{1 + e^{\beta_0^* + \underset{\sim}{\beta}' \underset{\sim}{x}}} = \frac{e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}{1 + e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}$$

$$\underset{\sim}{x}^* = \begin{pmatrix} 1 \\ \underset{\sim}{x} \end{pmatrix} ; \quad \underline{\beta}^* = \begin{pmatrix} \beta_0^* \\ \underset{\sim}{\beta} \end{pmatrix}$$

& $$p(\Pi_2 | \underset{\sim}{x}) = \frac{1}{1 + e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}.$$

$$f_Y(y_i) = P(Y = y_i) = \left(p(\Pi_1 | \underset{\sim}{x})\right)^{y_i} \left(1 - p(\Pi_1 | \underset{\sim}{x})\right)^{1 - y_i}$$

y_i is 0 or 1

Likelihood f^n

$$L(\underline{\beta}^*) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

$$L(\underline{\beta}^*) = \prod_{i=1}^{n} \left(p(\Pi_1 | \underset{\sim}{x}_i)\right)^{y_i} \left(1 - p(\Pi_1 | \underset{\sim}{x}_i)\right)^{1 - y_i}$$

log likelihood

$$\log L = \sum_{i=1}^{n} \left(y_i \log p(\Pi_1 | \underset{\sim}{x}_i) + (1 - y_i) \log \left(1 - p(\Pi_1 | \underset{\sim}{x}_i)\right) \right)$$

$$= \sum_{i=1}^{n} \left(y_i \log \left(\frac{p(\Pi_1 | \underset{\sim}{x}_i)}{1 - p(\Pi_1 | \underset{\sim}{x}_i)} \right) + \log \left(1 - p(\Pi_1 | \underset{\sim}{x}_i)\right) \right)$$

$$\log L = \sum_{i=1}^{n}\left(y_i\, \underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}_i^* - \log\left(1+e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}_i^*}\right)\right)$$

$$\frac{\partial \log L}{\partial \underset{\sim}{\beta}^*} = \sum_{i=1}^{n}\left(y_i \underset{\sim}{x}_i^* - \left(1+e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}_i^*}\right)^{-1} e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}_i^*}\, \underset{\sim}{x}_i^*\right)$$

$$= \sum_{i=1}^{n} \underset{\sim}{x}_i^* \left(y_i - \frac{e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}{1+e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}\right)$$

$$\frac{\partial \log L}{\partial \underset{\sim}{\beta}^*} = 0 \Rightarrow \sum_{i=1}^{n} \underset{\sim}{x}_i^* \left(y_i - \frac{e^{\underset{\sim}{\beta}^{*} \underset{\sim}{x}_i^*}}{1+e^{\underset{\sim}{\beta}^{*\prime} \underset{\sim}{x}^*}}\right) = 0 \quad \text{— } (*)$$

$(*)$ is a system of $p+1$ nonlinear equations which needs to be solved to ~~do~~ obtain the MLEs. An Iteratively ~~Reweighted~~ Reweighted Least Squares (IRLS) method is used for solving $(*)$ to get the MLEs.