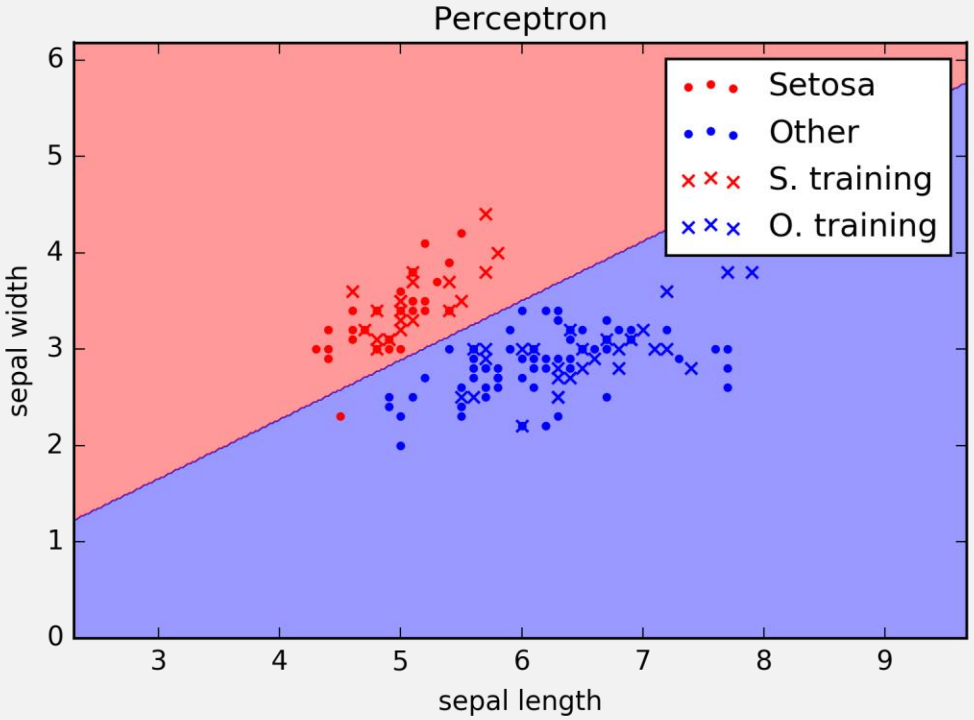
Classification

Linear Classification, Discriminant Function, Logistic Regression

Linear Classification

- Partition the input space into distinct regions, each labeled based on the classification.
- The boundaries of these regions can be either rough or smooth, depending on the prediction function.
- When these boundaries are linear, the method is referred to as linear classification.



Approaches of linear Classification

- Using Discriminant function
- Directly model a hyperplane

Linear Classification

- Consider the classification model or predictor $G(x)$ takes the set of discrete values $\mathcal{g}$.
- For linear classification, we are assume that the classifier $G(x)$ divides the input space into a collection of regions labelled according to the classification.
- The boundaries of this regions are called decision boundaries.
- These decision boundaries should be normal for linear classification.

Indicator Variable

- Each of the response categories are coded via an indicator variable. If target variable consists of K discrete values, then K indicator variables $\{Y_1, Y_2, \dots, Y_K\}$ represented for the target variable.
- $Y_k = 1$, if $\mathcal{g} = k$
- $Y_k = 0$, if $\mathcal{g} \neq k$
- Each row contains one 1 and remains 0

Discriminant Function

- Fit linear regression model for each indicator variable.
- Classify the largest fit.
- Let the fitted model for the k −th indicator variable is $f_k(x) = \hat{\beta}_{k0} + \hat{\beta}_{k1}^T x$
- The decision boundaries for the class k and l is the set of points for which $f_k(x) = f_l(x)$
- The hyperplane $\{x: (\hat{\beta}_{k0} - \hat{\beta}_{l0}) + (\hat{\beta}_{k1} - \hat{\beta}_{l1})^{Tx} = 0\}$
- This function is called discriminant function.

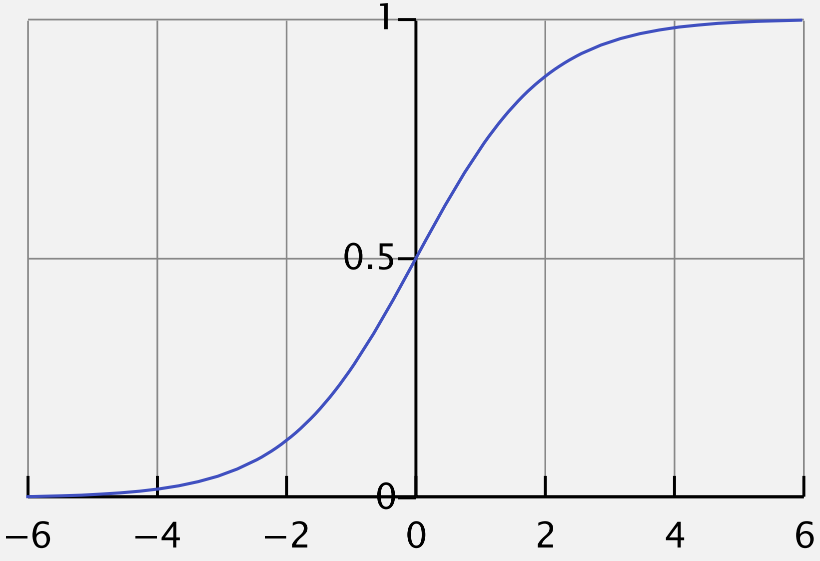
Linear Regression of Indicator Variables

- The target variable is a $N \times K$ matrix of 0 and 1. Fit the linear regression model to each columns of Y, and the fit is given by
- $\hat{Y} = X(X^T X)^{-1} X^T Y$
- The coefficient matrix $\mathcal{B}$ have $(d+1) \times K$ parameters.
- A new observation$x$ is classified as
  - ➢ Computed the fitted output$\hat{f}_k(x) = (1, x^T)\mathcal{B}$, a K vector
  - ➢ Identify the largest component and classify accordingly

$$G(x) = \arg\max_{k \in \mathcal{g}} \hat{f}_k(x)$$

Logistic Regression: Binary Classification

- Predicts the probability of occurring an event.
- $Input \rightarrow Probability$



Logistic Regression: Binary Classification

- Transformation required to get the probability
- Take $\log odds$ as linear regression
- $\log odds = \beta_0 + \beta_1 x$
- $odds$ is the ratio of probability of success and failure
- $p(x)$ is the probability of success, $1 - p(x)$ is the probability of failure

Logistic Regression: Binary Classification

- As log $odds$ will be expressed as linear regression $\log \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x$
- $p(x) = \left(1 - p(x)\right) \exp(\beta_0 + \beta_1 x)$
- $p(x) = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$
- The probability, that the instance x is in class 1and 0 is given below
- $P(g = 1|X = x) = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$
- $P(g = 0|X = x) = \frac{1}{1+\exp(\beta_0+\beta_1 x)}$

Multiple Logistic Regression

- For multiple classification for K class, same thing can be extended

- $P(g = k|X = x) = \frac{\exp(\beta_{k0}+\beta_{k1}^T x)}{1+\sum_{i=1}^{K} \exp(\beta_{i0}+\beta_{i1}^T x)}$

Logistic Regression: Binary Classification