Logistic Regression Model

Probabilistic Models for Classification

$$Pr(Y = y \mid X = x)$$

Logistic Regression:

Estimates $Pr(Y = y \mid X = x)$ directly.

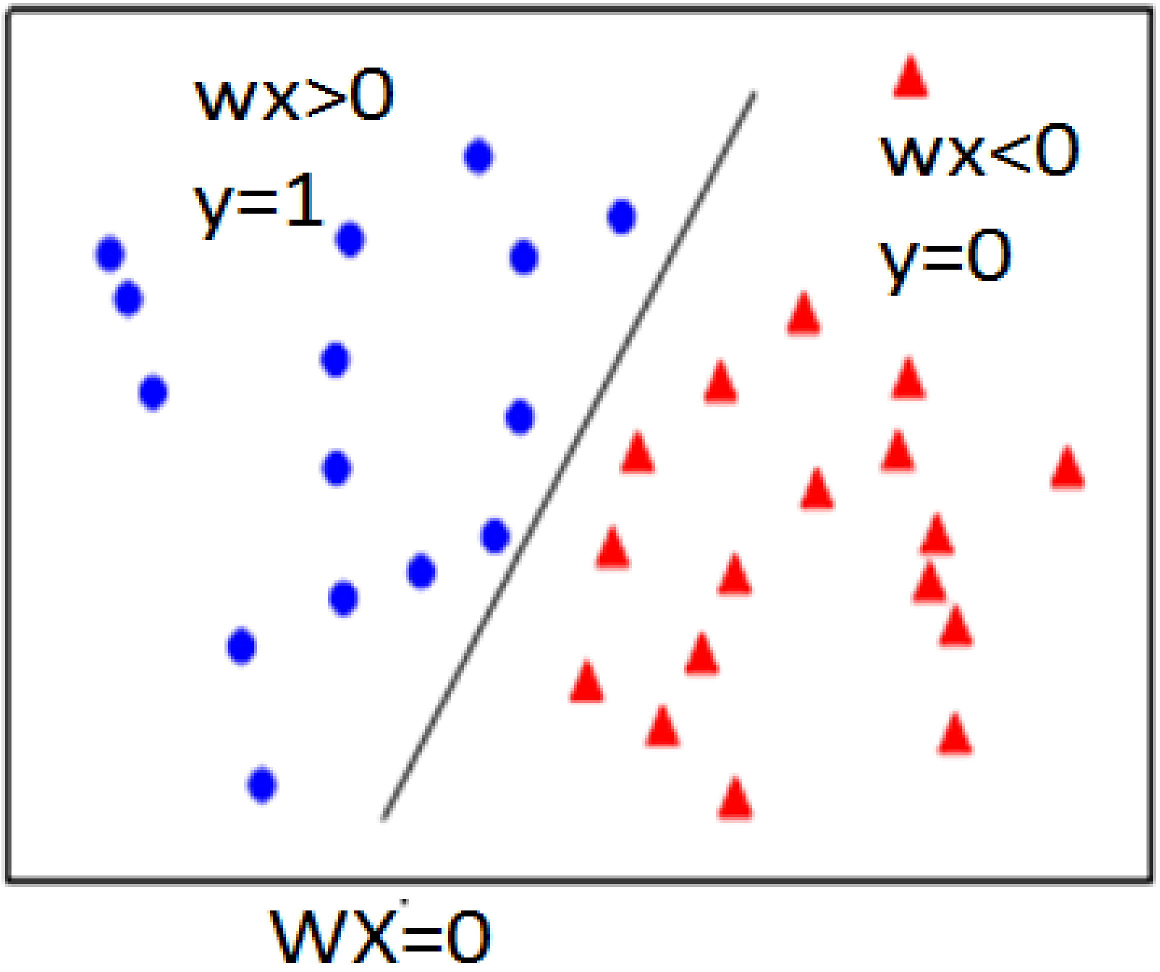
Naive Bayes:

Comes up with

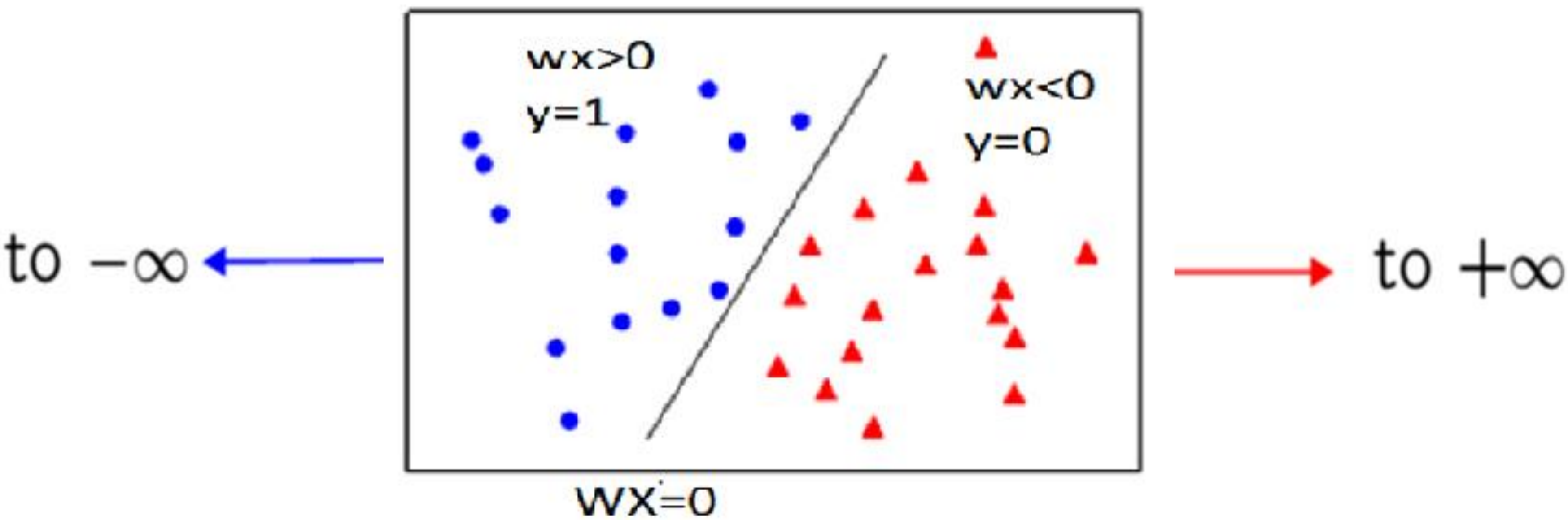
- $p(Y = y)$ (the marginal for Y)
- $P(X = x \mid Y = y)$ (the conditional distribution of X given Y)

and then uses Bayes Theorem to compute $P(Y = y \mid X = x)$.

Linear Descriminator



Linear models for classification



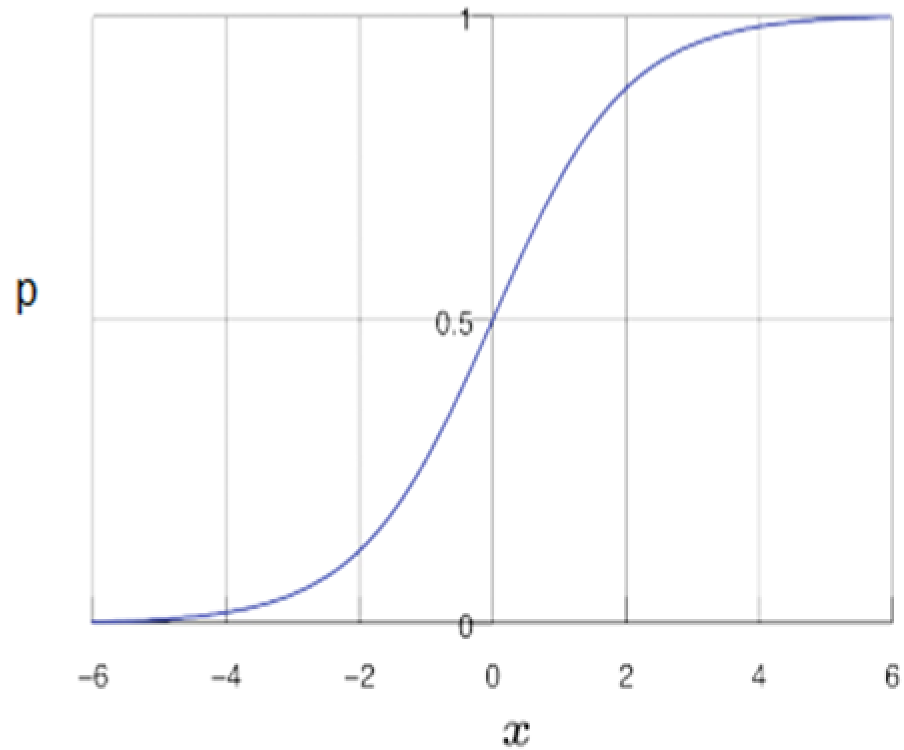
- the solution is to transform the parameter space so that

$$(-\infty, \infty) \rightarrow (0,1)$$

Link function: Logistic Function

$$h\ (x) = \frac{1}{1+e^{-\ x}}$$

- **If** $h\ (x) \geq 0.5$ **or equivalently** $x \geq 0$ **predict** y = 1
- **If** $h\ (x) < 0.5$ **or equivalently** $x < 0$ **predict** y = 0



Generalized Linear Models

All generalized linear models have the following three characteristics:

1. A probability distribution describing the outcome variable
2. A linear model
   - $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$
3. A link function that relates the linear model to the parameter of the outcome distribution
   - $g(p) = \eta$ or $p = g^{-1}(\eta)$

Model Learning: Parameter Estimation

- Use Maximum Likelihood Estimation Technique for estimating parameters of Logistic Regression Model

- That means, we choose the parameter values that make the data we have seen most likely.

Maximum Likelihood Estimation

Assume

$$
\begin{aligned}
p(y = 1|\mathbf{x}; \mathbf{w}) &= \sigma(\mathbf{w}^\top \mathbf{x}) \\
p(y = 0|\mathbf{x}; \mathbf{w}) &= 1 - \sigma(\mathbf{w}^\top \mathbf{x})
\end{aligned}
$$

write this more compactly as

$$
p(y|\mathbf{x}; \mathbf{w}) = \left(\sigma(\mathbf{w}^\top \mathbf{x})\right)^y \left(1 - \sigma(\mathbf{w}^\top \mathbf{x})\right)^{(1-y)}
$$

Then the likelihood (assuming independence) is

$$
p(\mathbf{y}|\mathbf{x}; \mathbf{w}) \sim \prod_i^n \left(\sigma(\mathbf{w}^\top \mathbf{x}_i)\right)^{y_i} \left(1 - \sigma(\mathbf{w}^\top \mathbf{x}_i)\right)^{(1-y_i)}
$$

and the negative log likelihood is

$$
L(\mathbf{w}) = -\sum_i^n y_i \log \sigma(\mathbf{w}^\top \mathbf{x}_i) + (1 - y_i) \log(1 - \sigma(\mathbf{w}^\top \mathbf{x}_i))
$$

Objective Function: Log Loss

Because we always have *y* = *0* or *y* = *1* we can simplify the cost function definition to

$$Cost(h\ (x), y) \ = \ -y\ log(h\ (x)) - (1-y)log(1-h\ (x))$$

Log Loss function applied for Logistic Regression: Pass the value of WX for h

$$L(\mathbf{w}) = -\sum_{i}^{n} y_i \log h(\mathbf{w}\ \ \mathbf{x}_i) + (1-y_i)\log(1-\ h(\mathbf{w}\ \ \mathbf{x}_i))$$

Minimize L(w).

Solving optimization with GD

$$\frac{\partial}{\partial w_j} L(\mathbf{w}) = -\sum_i \left(y_i - \mathrm{h}(\mathbf{w}^\top \mathbf{x}_i) \right) x_j$$

Implement Batch GD and Stochastic GD solutions for the objective function using above gradient.