Logistic Regression

- Two classes
- Model

$$p(1|\underline{x}) = \frac{\exp(\beta_0 + \underline{\beta}^T \underline{x})}{1 + \exp(\beta_0 + \underline{\beta}^T \underline{x})}$$

$$p(-1|\underline{x}) = \frac{1}{1 + \exp(\beta_0 + \underline{\beta}^T \underline{x})} =$$

$$g(t) = \frac{e^t}{1 + e^t} \quad \text{is}$$

Estimating Parameters

Given training data $\mathcal{T} = \{(\underline{x}_i, y_i)\}_{i=1}^{N}$

find β_0 and $\underline{\beta}$ that maximize

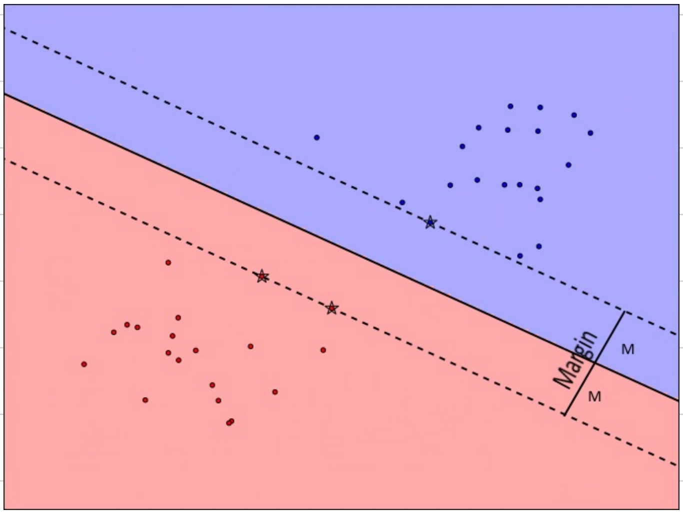
$$h(\beta_0, \underline{\beta}) =$$

Logistic Regression is a

$$p(1|\underline{x}) \underset{-1}{\overset{1}{\gtrless}} p(-1|\underline{x}) \equiv$$

Support Vector Machine (SVM)

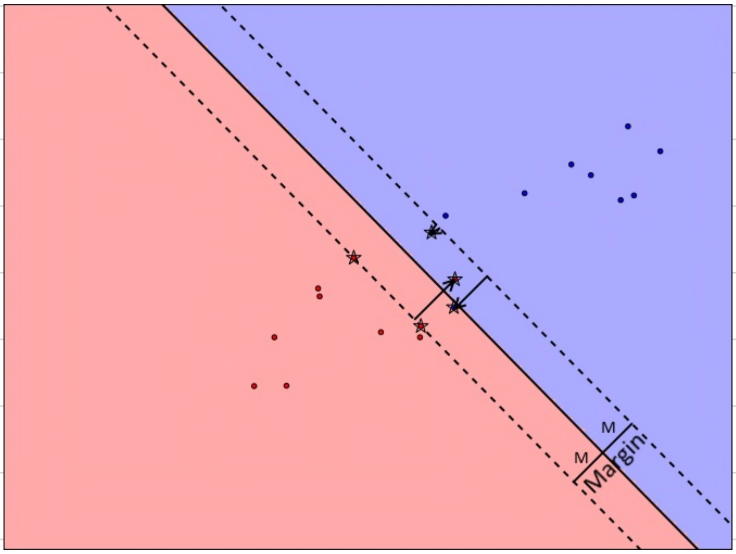
Linearly Separable Training Data



Infinitely many linear boundaries separating classes

SVM maximizes

Training Data Not Linearly Separable



1. Fix decision boundary and margin
2. Compute distances of points on "wrong" side of margin to margin boundary
3. Compute penalty
4. Maximize m subject to
5. Choose boundary to

Naive Bayes Classifier

- Recall $\hat{y}_{Bayes} = \arg\max_{y=1,\dots,M} \pi_y \, p(\underline{x} | y)$

  Feature vector $\underline{x} = [x_1 \; x_2 \cdots x_d]^T$

- Naive Bayes assumes that $x_1, x_2, \dots, x_d$ are

  $p(\underline{x} | y) =$

- Get estimates

  $\hat{y}_{NB} =$

  $=$

Pros and Cons of NB Classifier

+ estimating $p(\underline{x} | y)$

  estimating $p(x_1|y), p(x_2|y), \dots p(x_d|y)$ separately

+ Can mix

− independence assumption is

Finding $\hat{p}(x_j | y)$ using training data

Convenient to use parametric model:

$$p(x_j | y) = p(x_j | y, \theta_{jy})$$

Then, $\hat{p}(x_j | y) =$

Example (continuous x_j)

$$p(x_j | y, \theta_{jy}) = \frac{1}{\sqrt{2\pi \sigma_{jy}^2}} e^{-\frac{(x_j - \mu_{jy})^2}{2\sigma_{jy}^2}} \leftarrow$$

$\theta_{jy} =$

Given training data $\{(\underline{x}_i, y_i)\}_{i=1}^{N}$,

$\hat{\mu}_{jy} =$, $\hat{\sigma}_{jy}^2 =$

Example (discrete or categorical x_j)

x_j takes values in set $\{a_1, a_2, \ldots, a_\ell\}$ with probabilities $\rho_{1,jy}, \rho_{2,jy}, \ldots \rho_{\ell jy}$, for class y.

$\theta_{jy} = (\rho_{1,jy}, \rho_{2,jy}, \ldots \rho_{\ell,jy})$, $\hat{\theta}_{jy} = (\hat{\rho}_{1,jy}, \hat{\rho}_{2,jy}, \ldots \tilde{\rho}_{\ell,jy})$

$\hat{\rho}_{k,jy} =$