Lecture 4

Chiranjib Bhattacharyya

chibha@chalmers.se

January 27, 2012

Agenda

- Fisher Discriminant
- Perceptron

Iris Flower Dataset

Versicolor

Setosa

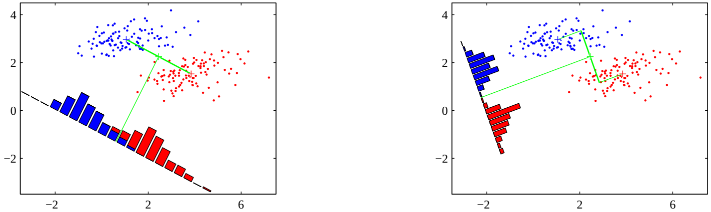
Virginica

Features: Sepal Length, Sepal Width, Petal Length, Petal Width

150 observations, 3 classes

Fisher Discriminant

Source: Pattern Recognition and Machine Learning (Chris Bishop)



Projection

Projection of $x \in \mathbb{R}^d$ on $w \in \mathbb{R}^d$ is

$$r_w(x) = \|x\| cos\theta \;\; cos\theta = \frac{w^\top x}{\|x\|\|w\|}$$

Let $E(x) = \mu$, and $\Sigma = E(x-\mu)(x-\mu)^\top$

$$E(r_w(x)) = \frac{w^\top \mu}{\|w\|} \;\; E\left(r_w(x) - E(r_w(x)\right)^2 = \frac{w^\top \Sigma w}{\|w\|^2}$$

Fisher Discriminant

Let (μ_1, Σ_1) be the mean and covariance of class 1 and (μ_2, Σ_2) be the mean and covariance of class 2.

$$J(w) = max_w \frac{\left(w^\top(\mu_1 - \mu_2)\right)^2}{w^\top S w}$$

$w = S^{-1}(\mu_1 - \mu_2)$ $S = \Sigma_1 + \Sigma_2$

Linear Classifiers

Let us a define a linear classifier w, b on $D = \{(x_i, y_i) | i = 1, \dots, n\}$ such that

$$y_i = sign(w^\top x_i + b) \;\; i = 1, \dots, n$$

Linear Classifiers

Let us a define a linear classifier w, b on $D = \{(x_i, y_i) | i = 1, \ldots, n\}$ such that

$$y_i = sign(w^\top x_i + b) \;\; i = 1, \ldots, n$$

equivalent to

$$y_i(w^\top x_i + b) \geq 0 \;\; i = 1, \ldots, n$$

Perceptron Algorithm

Let $\tilde{w} = [w^\top b]^\top \;\; \tilde{x} = [x^\top 1]^\top$.
mistake happens if $y \neq sign(\tilde{w}^\top \tilde{x})$

Perceptron Algorithm

Set $\tilde{w}_1 = 0$, $t = 0$
For each example $x \in D$
check if $\tilde{w}(t)$ makes a mistake on x. then update
$\tilde{w}(t+1) = \tilde{w}(t) + y_t \tilde{x}$, and $t \rightarrow t+1$
Continue till there are no mistakes on the training set

Perceptron Algorithm

Theorem

Let $D = \{(x_i, y_i) | i = 1, \ldots, n\}$ and $\|x_i\| = 1$ $i = 1, \ldots, n$. If there exists $\tilde{w}^*, \gamma$ such that

$$\gamma = min_{x_i \in D} y_i \tilde{x}_i^\top \tilde{w}^*$$

then the perceptron algorithm converges after making atmost $\frac{1}{\gamma^2}$mistakes

Proof Sketch: After every update

$$\tilde{w}_{t+1}^\top \tilde{w}^* \geq \tilde{w}_t^\top \tilde{w}^* + \gamma \quad \|\tilde{w}_{t+1}\|^2 \leq \|\tilde{w}_t\|^2 + 1$$

Perceptron Algorithm

Proof:
if T is the number of mistakes

$$\tilde{w}_{T+1}^{\top}\tilde{w}^{*} \geq T\gamma \quad \|\tilde{w}_{T+1}\| \leq \sqrt{T}$$

For any two vectors $u^{\top}v \leq \|u\|\|v\|$

$$T\gamma \leq \tilde{w}_{T+1}^{\top}\tilde{w}^{*} \leq \|\tilde{w}_{T+1}\| \leq \sqrt{T}$$

and the claim follows