# Linear Discriminants

## CSci 5525: Machine Learning

Instructor: Paul Schrater

# Discriminant Functions

- One of the simplest representation for a 2-class problem

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

- Class assignment based on $\text{sign}(f(\mathbf{x}))$
  - If $f(\mathbf{x}) \geq 0$, $\text{sign}(f(\mathbf{x})) = +1$, then $\mathbf{x} \in C_1$, otherwise $\mathbf{x} \in C_2$
- $\mathbf{w}$ is orthogonal to the decision boundary $f(\mathbf{x}) = w_0$
- With $\tilde{\mathbf{w}} = (\mathbf{w}, w_0)$ and $\tilde{\mathbf{x}} = (\mathbf{x}, 1)$, we have

$$f(\mathbf{x}) = \tilde{\mathbf{w}}^T \tilde{\mathbf{x}}$$

- At times, we will ignore the offset term $w_0$ w.l.o.g. (without loss of generality)

# Least Squares for Classification

- Consider a training dataset $\{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^{N}$ for a $K$-class problem
  - $\mathbf{y}_n$ encodes the class membership, say $\mathbf{y}_n^T = (0, 1, 0, 0)$
  - $Y$ : $N \times K$ matrix with rows $\mathbf{y}_n^T$
  - $X$ : $N \times D$ matrix with rows $\mathbf{x}_n^T$
  - $W$ : $D \times K$ matrix with columns $\mathbf{w}_k$
  - Goal:

$$\mathbf{w}_k^T \mathbf{x}_n = \mathbf{x}_n^T \mathbf{w}_k = X_{n,:}\mathbf{w}_k \approx Y_{nk}$$

- The sum-of-squares error to be minimized over $W$ is

$$E(W) = \frac{1}{2}\sum_{k=1}^{K}\sum_{n=1}^{N} \| Y_{nk} - X_{n,:}\mathbf{w}_k \|^2 = \frac{1}{2}\,\mathrm{Tr}\left\{(Y - XW)^T(Y - XW)\right\}$$

# Least Squares for Classification (Contd.)

- The sum-of-squares error to be minimized over $W$ is

$$E(W) = \frac{1}{2}\sum_{k=1}^{K}\sum_{n=1}^{N}\|Y_{nk} - X_{n,:}\mathbf{w}_k\|^2 \; = \; \frac{1}{2}\mathrm{Tr}\left\{(Y - XW)^T(Y - XW)\right\}$$

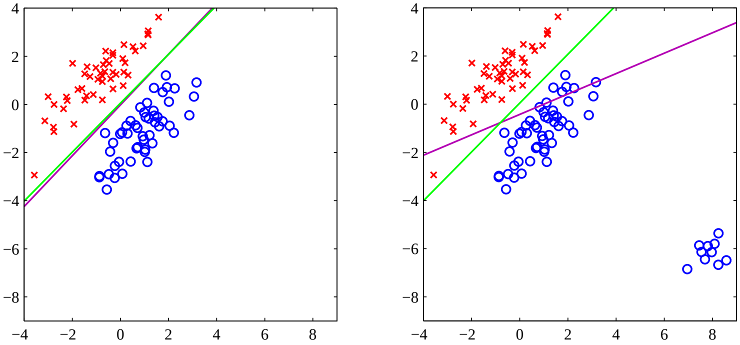
- The problem has a closed form solution

$$W = (X^T X)^{-1} X^T Y = X^{\dagger} Y$$

- Solving each problem separately: $\mathbf{w}_k = X^{\dagger}\mathbf{y}_k$
- The discriminant function has the following form

$$f(\mathbf{x}) = W^T\mathbf{x} = Y^T(X^{\dagger})^T\mathbf{x}$$

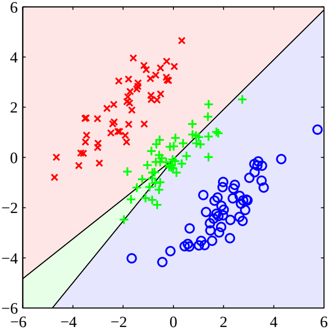
# Least Squares is Noise Sensitive



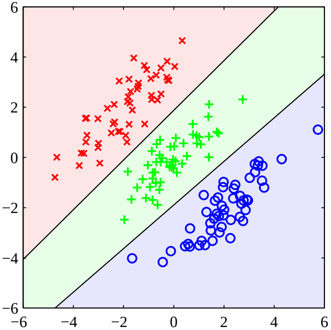

Least Squares vs Logistic

Least Squares with noise

# Least Squares for Multiclass Problems




Least Squares




Logistic Regression

# Classification by Projection

- Classify after dimensionality reduction
  - Project $D$ dimensional data $\mathbf{x}$ to 1 dimensions: $\mathbf{w}^T\mathbf{x}$
  - Make sure class separation is maximized
- If $\mathbf{m}_1, \mathbf{m}_2$ are the means of the two classes

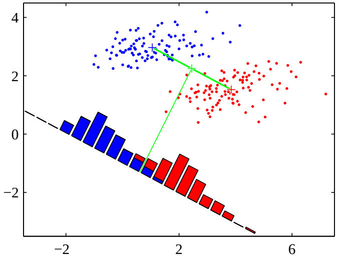
$$\max_{\|\mathbf{w}\|^2=1} \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)$$

- Performing the optimization (using 'Langrange multipliers')

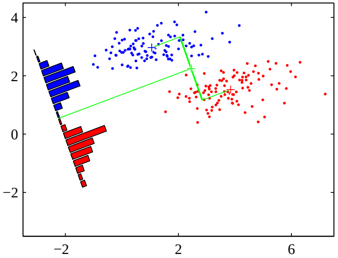
$$\mathbf{w} \propto (\mathbf{m}_2 - \mathbf{m}_1)$$

- May be problematic if data has non-diagonal covariance

# Classification by Projection (Contd.)




Classification by Projection




Fisher's Linear Discriminant

# Fisher's Linear Discriminant

- Desirable to have low within class variance

$$\sigma_k^2 = \sum_{\mathbf{x}_n \in C_k} \|w^T(\mathbf{x}_n - \mathbf{m}_k)\|^2$$

- Between-class and within-class covariance matrices

$$S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T$$
$$S_w = \sum_{\mathbf{x}_n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^T + \sum_{\mathbf{x}_n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^T$$

- Fisher's criterion: Ratio of between-class and within-class variance

$$J(\mathbf{w}) = \frac{\|\mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)\|^2}{\sigma_1^2 + \sigma_2^2} = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

# Fisher's Linear Discriminant (Contd.)

- Fisher's criterion is

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

- A 'direct calculation' gives

$$\mathbf{w} \propto S_w^{-1}(\mathbf{m}_2 - \mathbf{m}_1)$$

- A linear discriminant can be constructed using $\mathbf{w}$
  - Construct the projected version of the data $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$
  - Choose a threshold $w_0$ to form linear discriminant $f(\mathbf{x}) \geq w_0$
- Extension to multiclass: Project to $(K-1)$ dimensions
- Need to train a classifier in the low dimensional representation