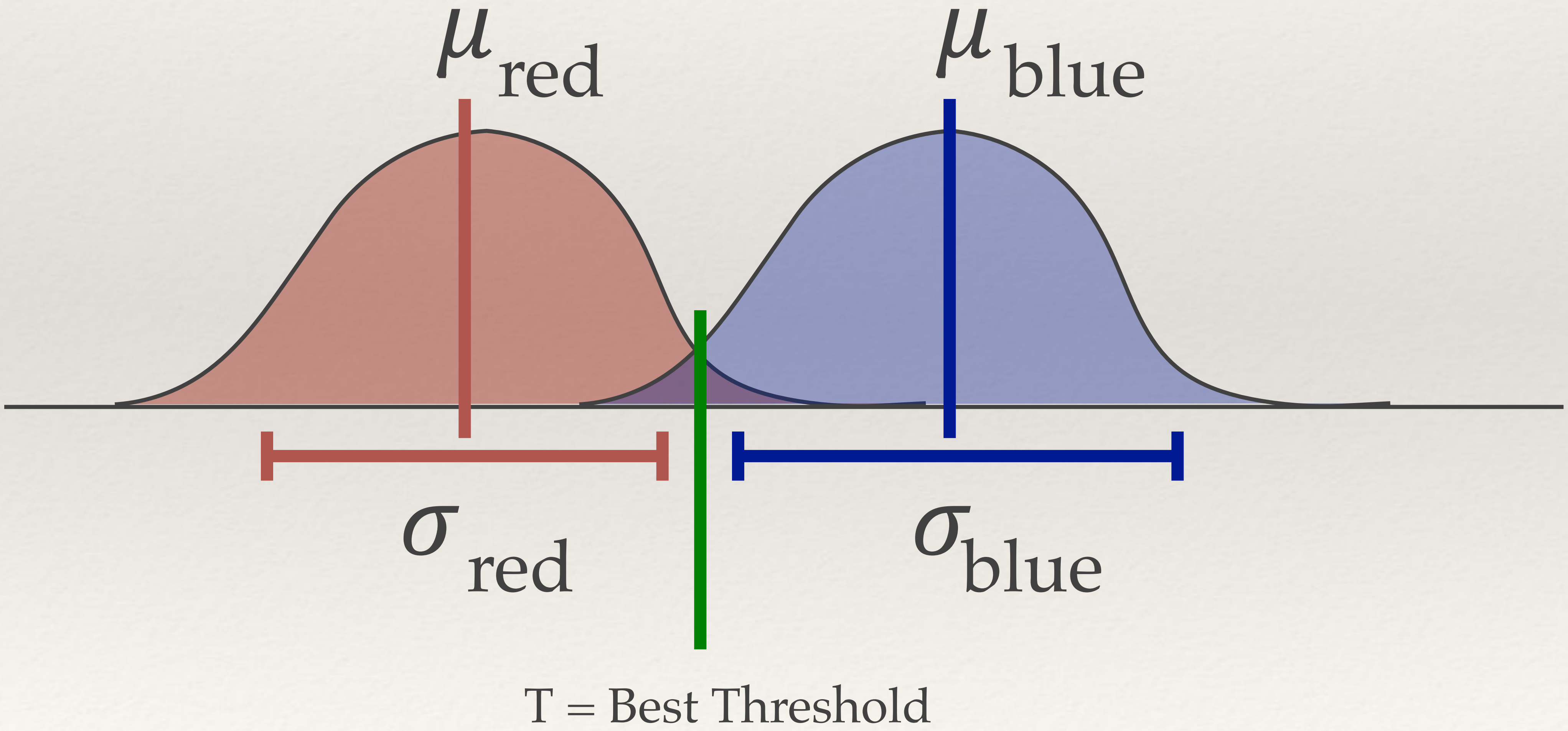
Michael Grossberg

Intro to Data Science CS59969

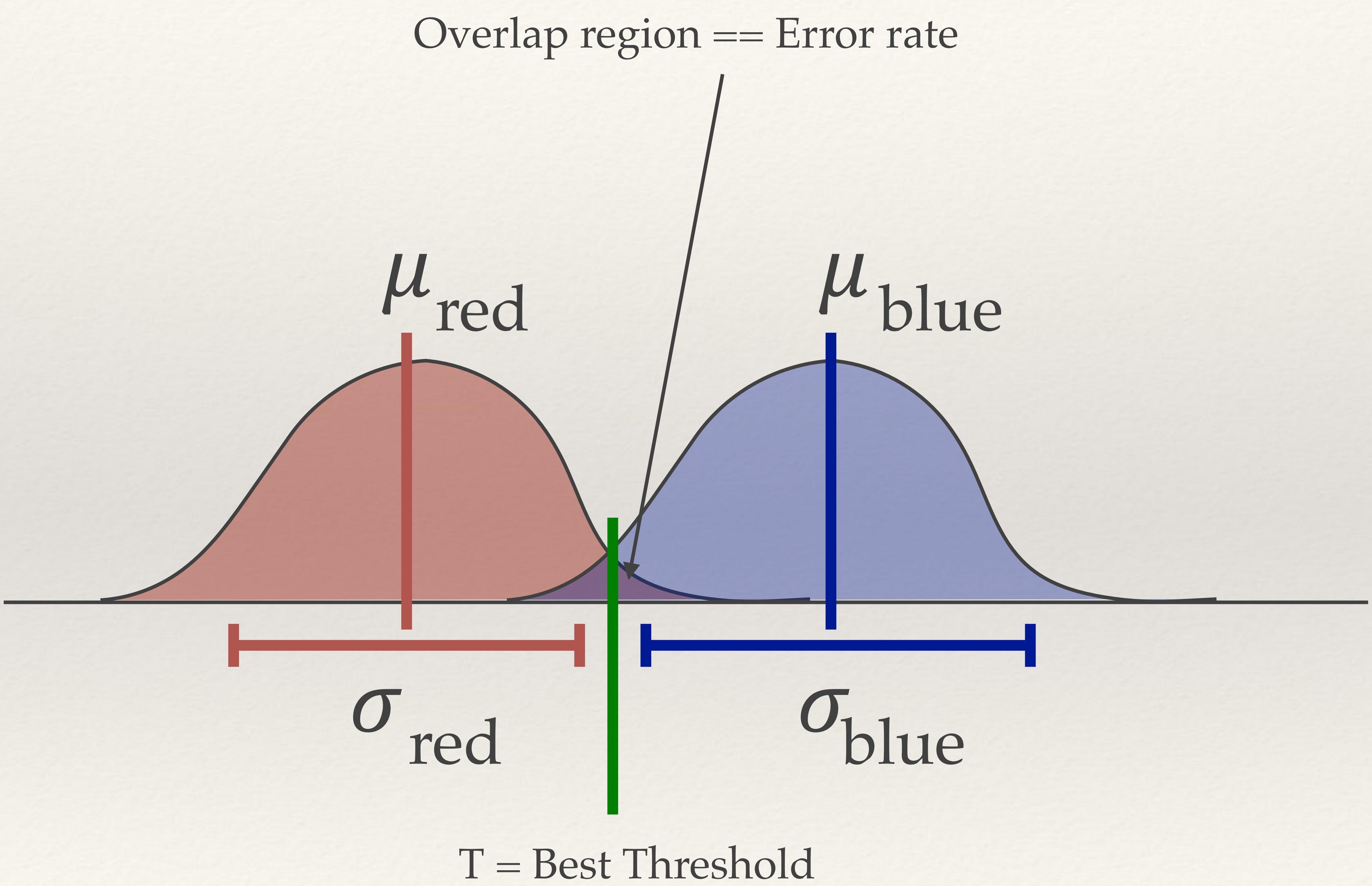
LDA

Linear Classifier

1-D Two Class



1-D Two Class

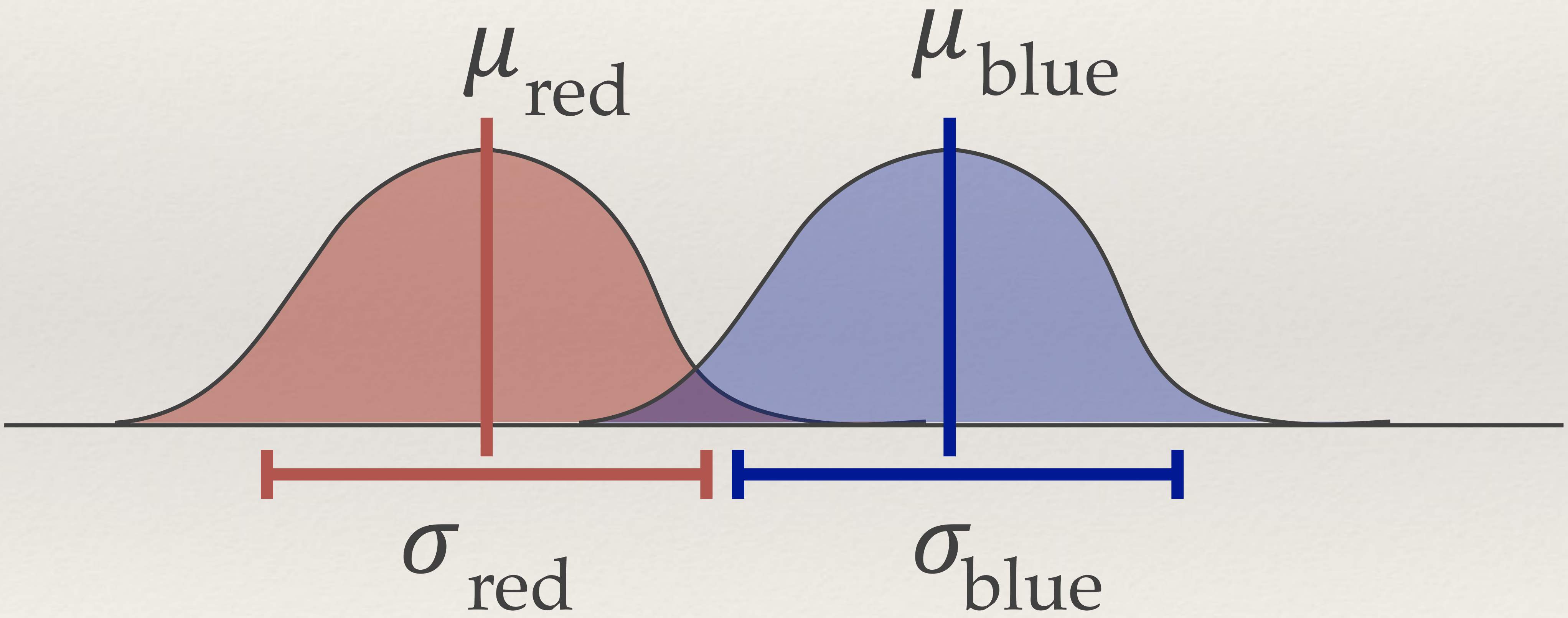


Overlap region == Error rate
μ_{red}
μ_{blue}
σ_{red}
σ_{blue}
T = Best Threshold

1-D Two Class

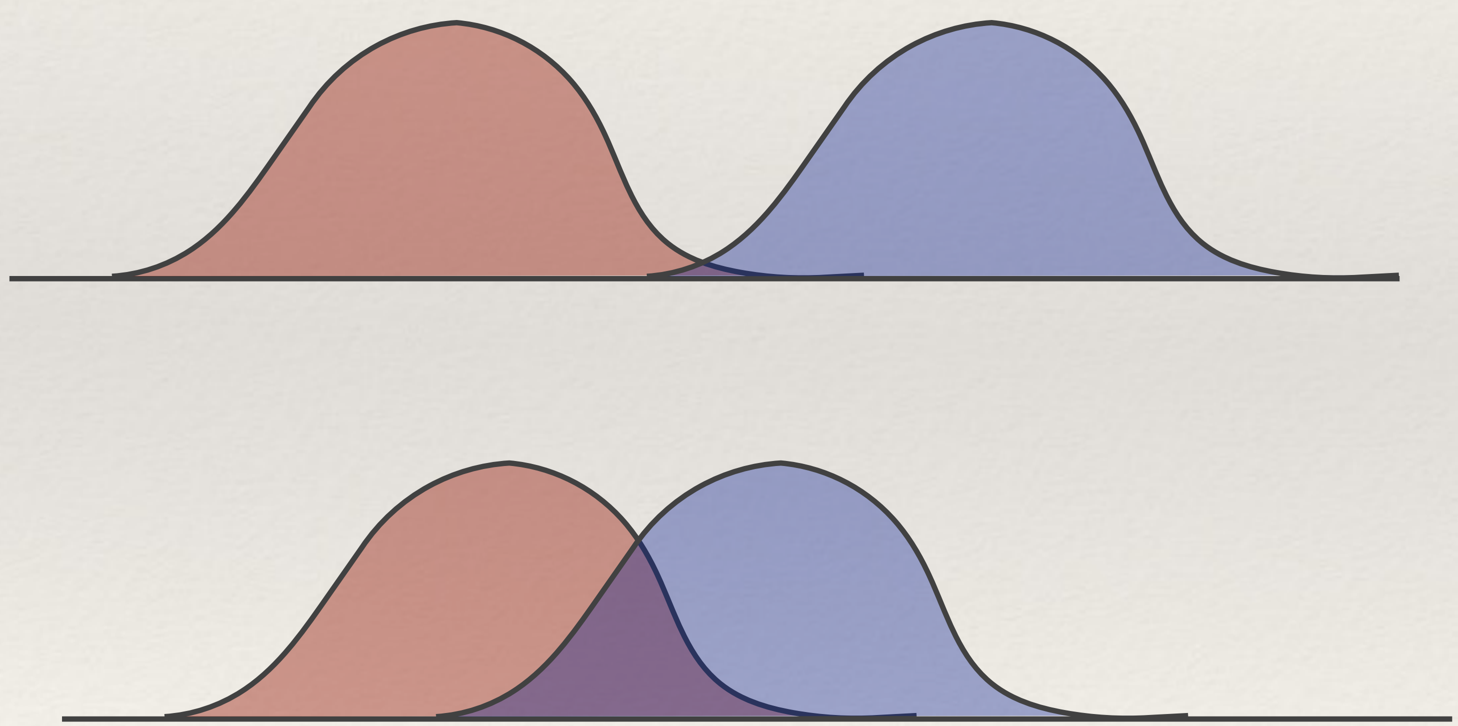
$$\sigma^2_{between} = \mu_{red} - \mu_{blue}$$

$$\sigma^2_{within} = 1/2(\sigma^2_{red} + \sigma^2_{blue})$$



Good/Bad Separation

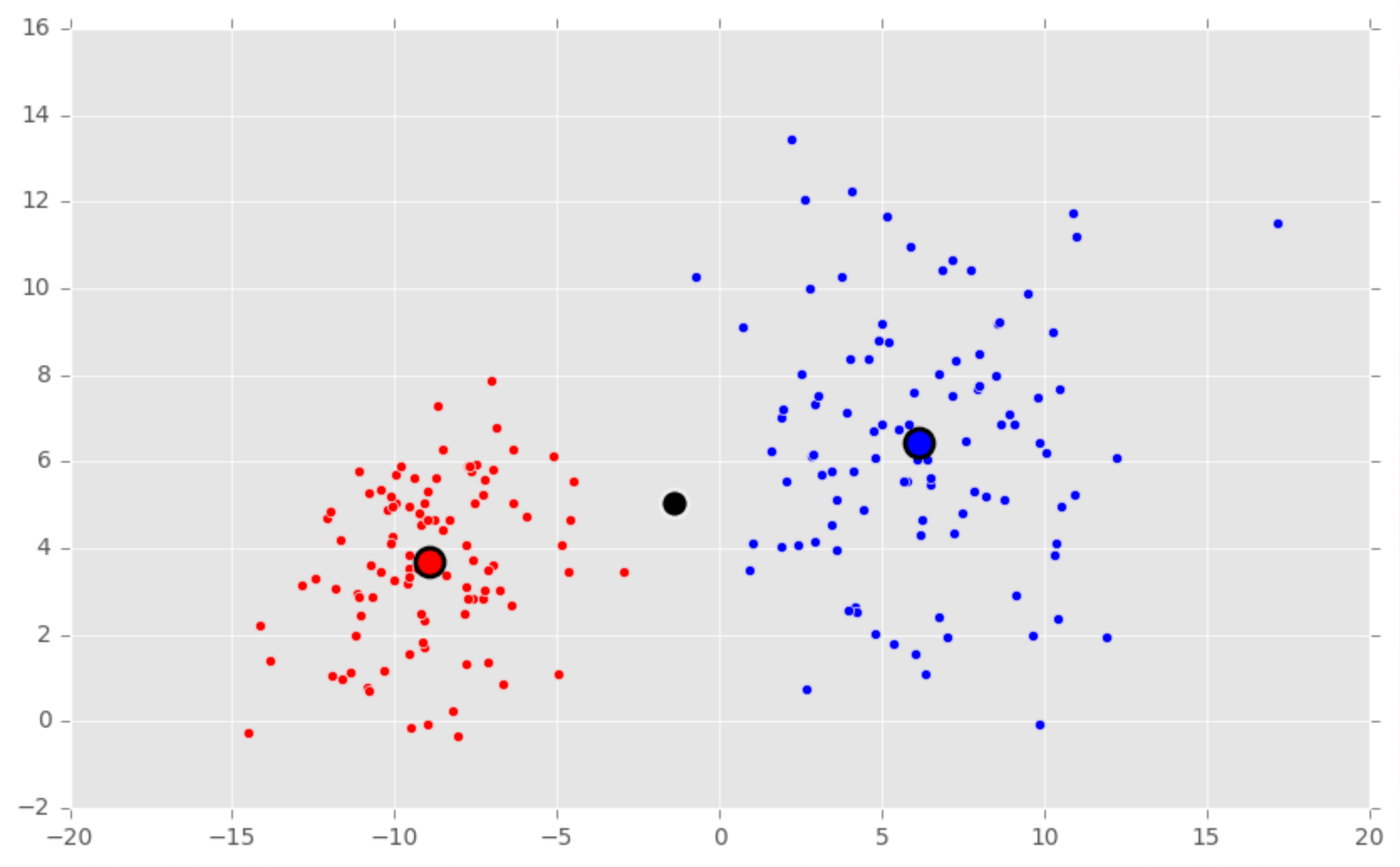
$$S = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}}$$



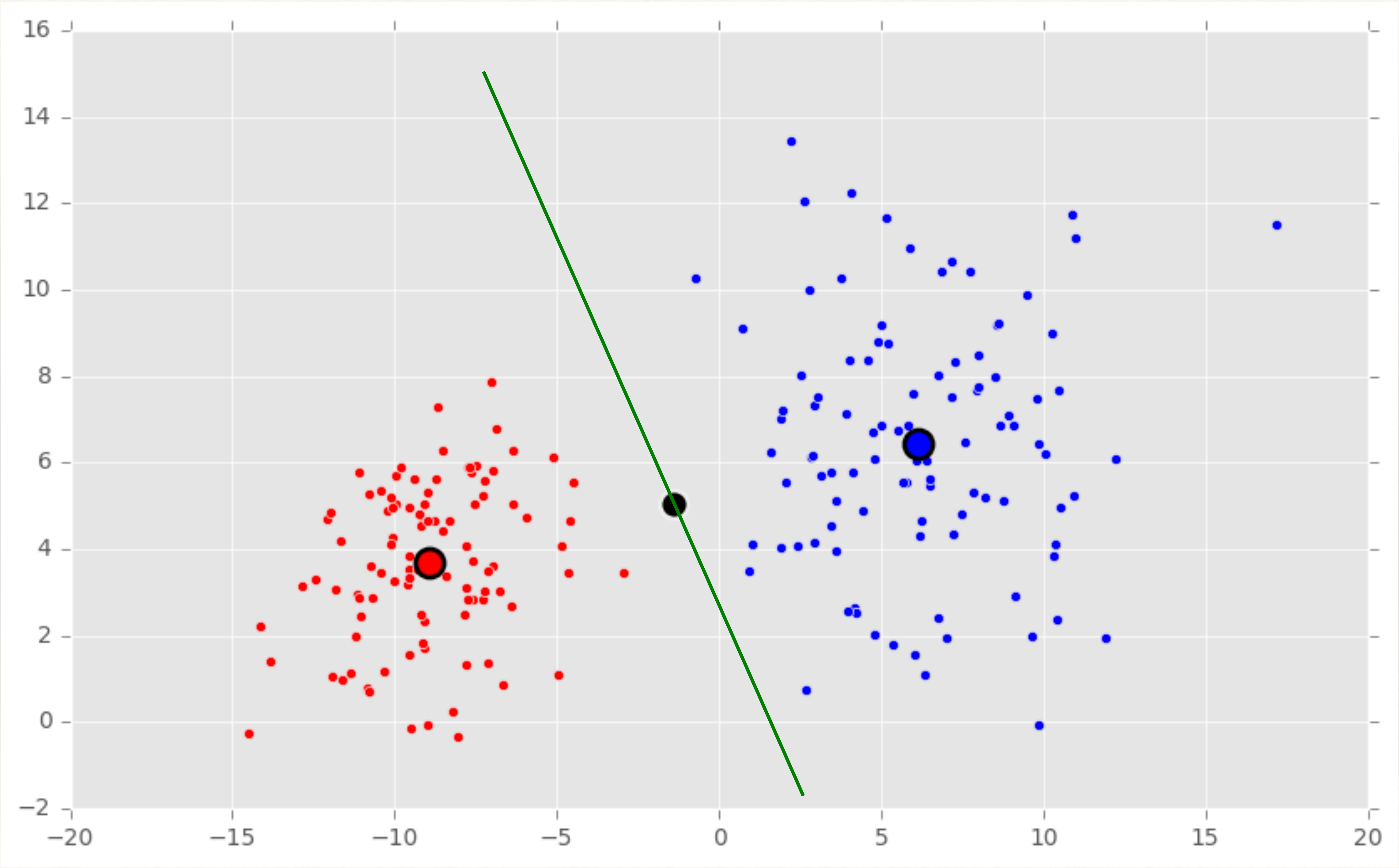
$S \uparrow$ Good

$S \downarrow$ Bad

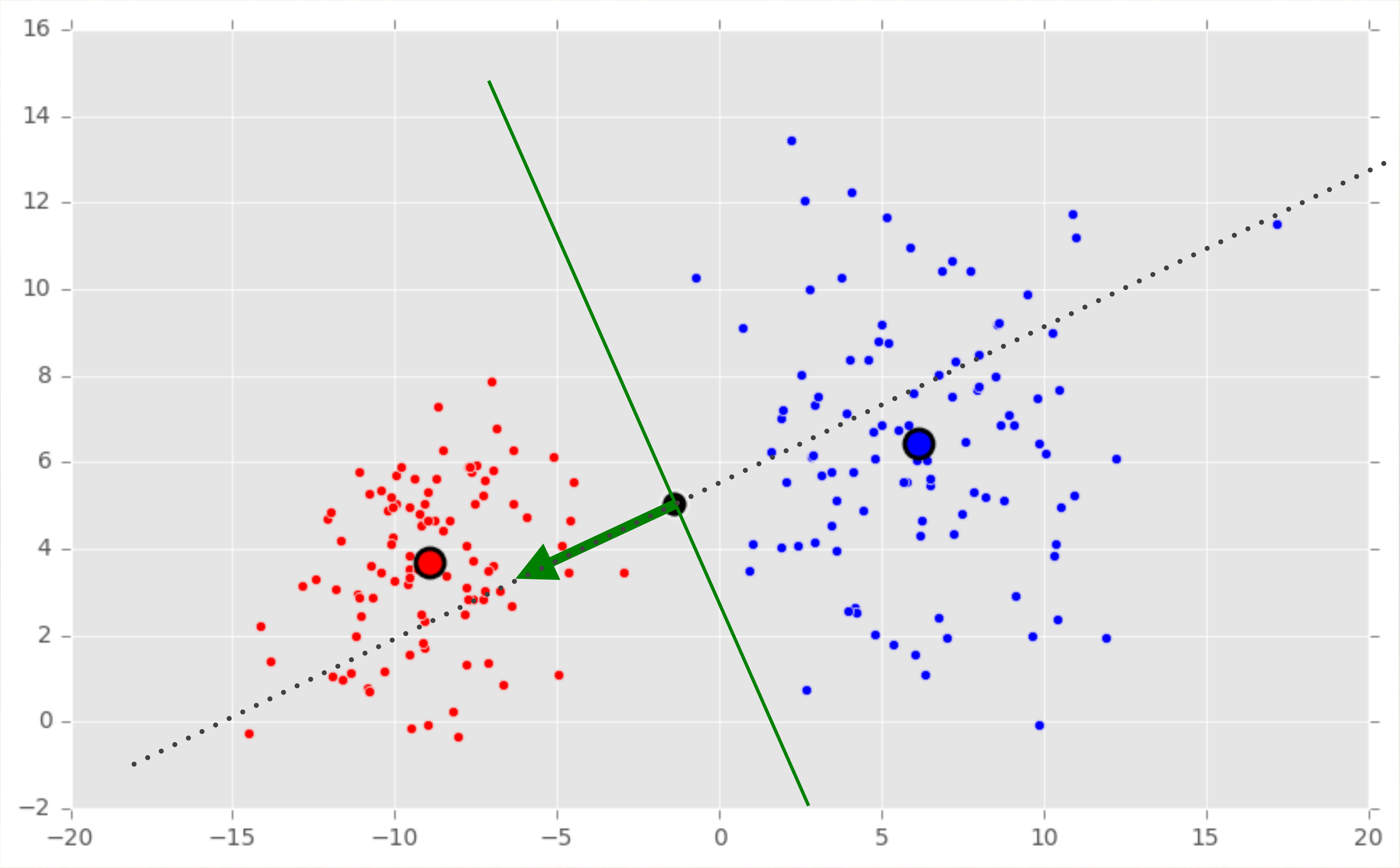
Binary Classification Problem



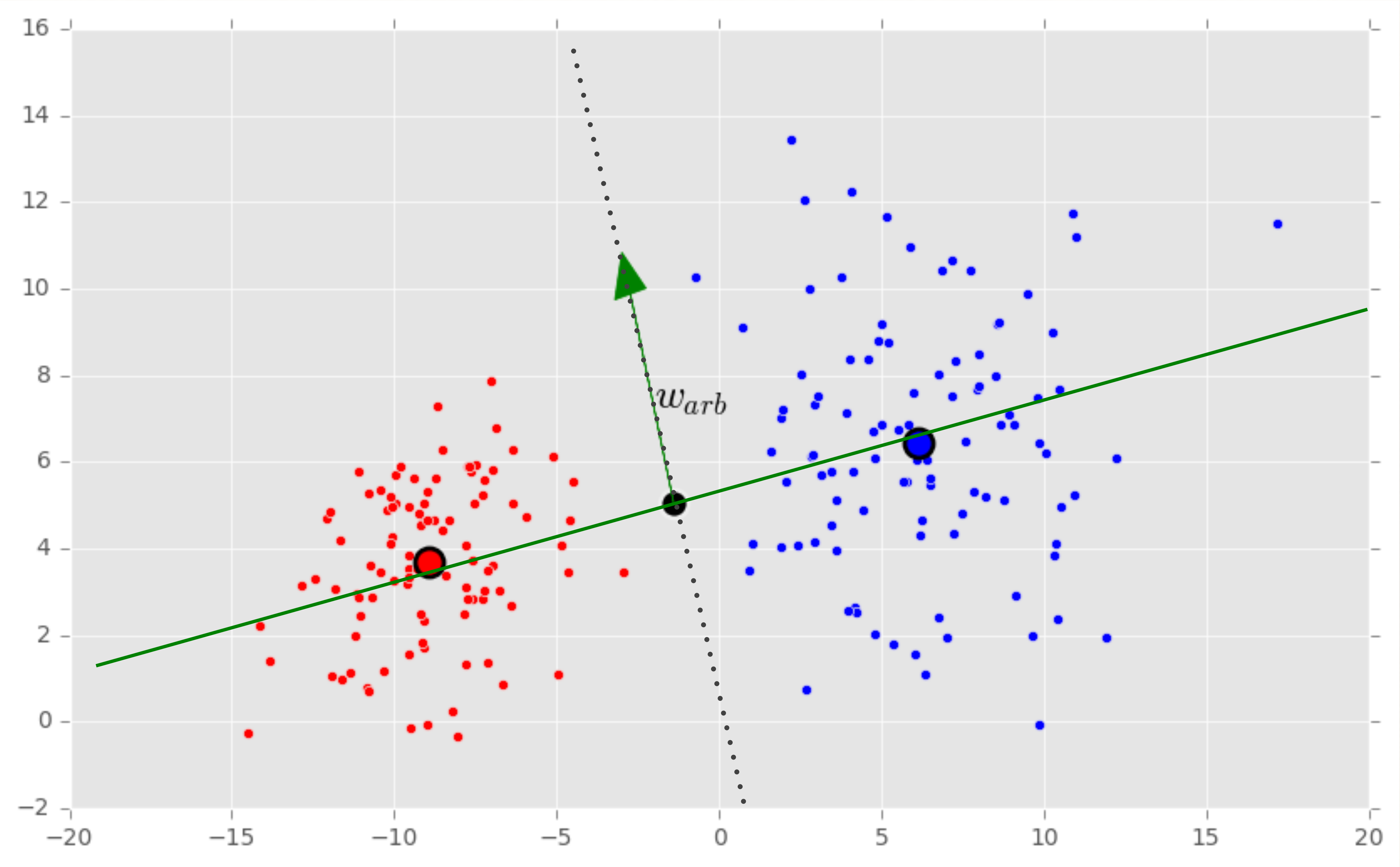
Finding Linear Separator



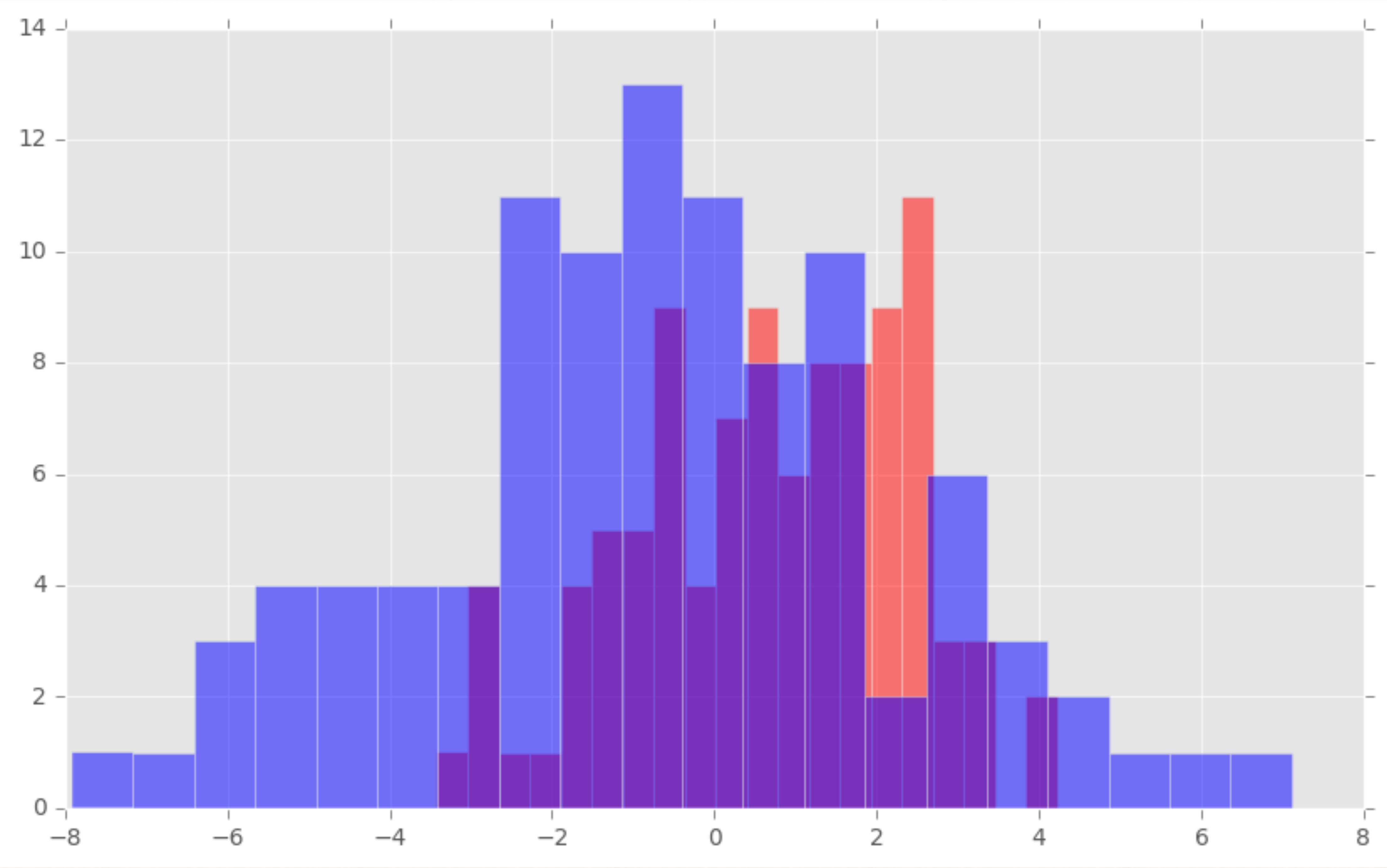
Equivalent: Find good projection



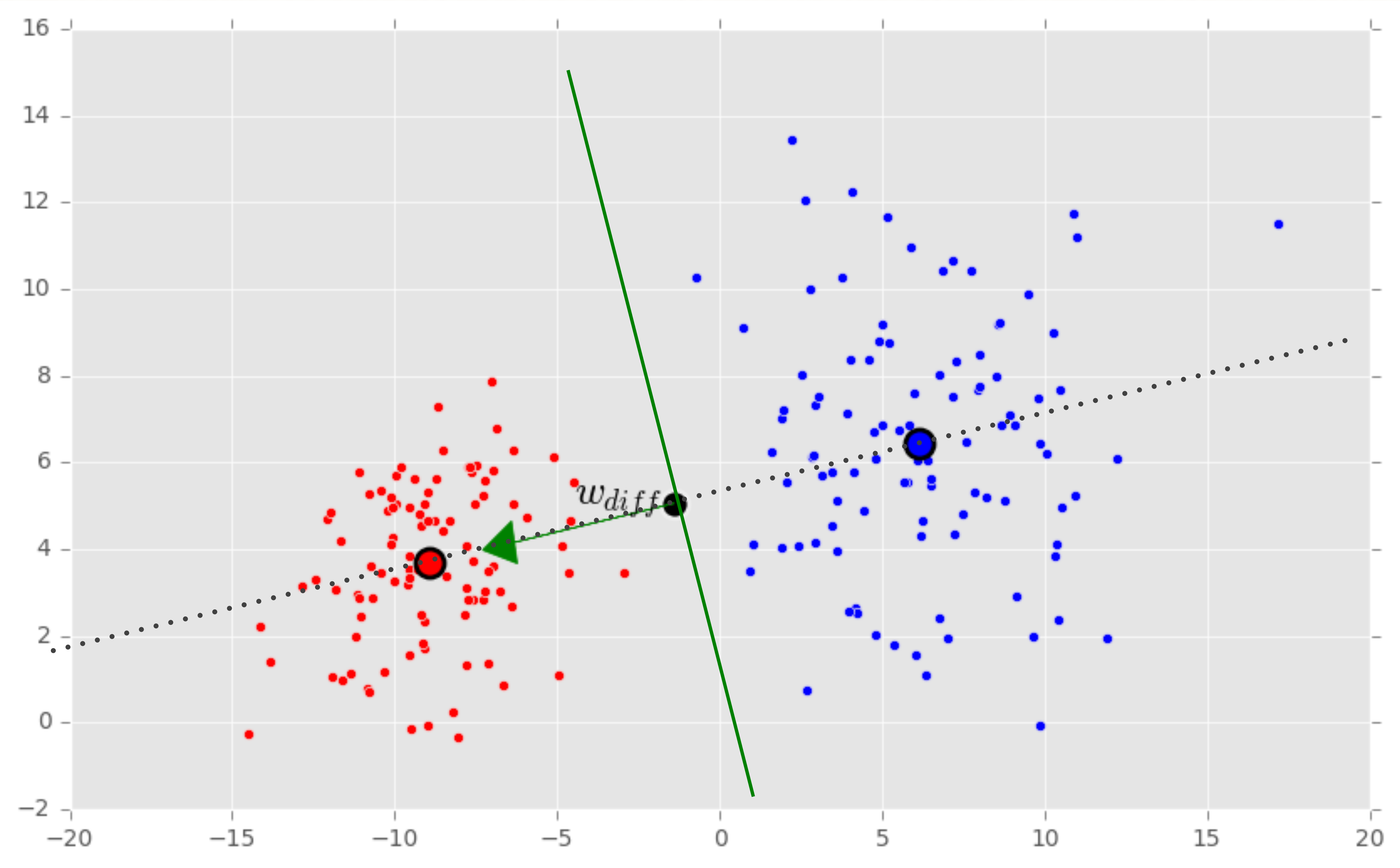
Example poor projection



Histogram of projection



Better Projection (difference)



Histogram of projection

More Generally

$\vec{\mu}_0, \vec{\mu}_1$ Class Means

Σ_0, Σ_1 Covariance Matrices

$\vec{w} \cdot \vec{x}$ w Projection "feature"

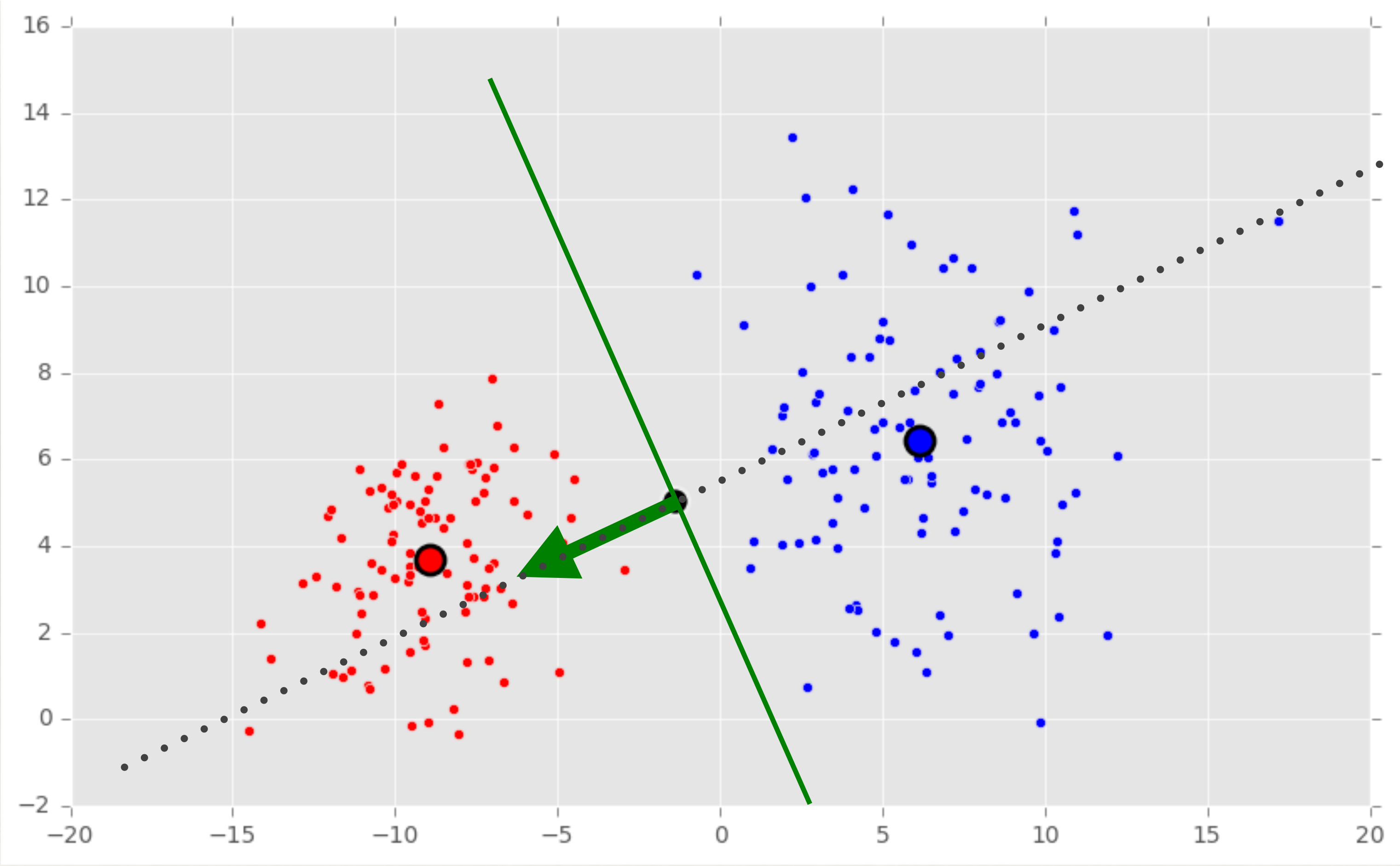
$$S = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{(\vec{w} \cdot \vec{\mu}_1 - \vec{w} \cdot \vec{\mu}_0)^2}{\vec{w}^T \Sigma_1 \vec{w} + \vec{w}^T \Sigma_0 \vec{w}} = \frac{(\vec{w} \cdot (\vec{\mu}_1 - \vec{\mu}_0))^2}{\vec{w}^T (\Sigma_0 + \Sigma_1) \vec{w}}$$

$$\vec{w} \propto (\Sigma_0 + \Sigma_1)^{-1} (\vec{\mu}_1 - \vec{\mu}_0)$$

Maximum Separation

Hyper-Plane Decision Boundary

$$c = \vec{w} \cdot \frac{1}{2}(\vec{\mu}_0 + \vec{\mu}_1) = \frac{1}{2}\vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1 - \frac{1}{2}\vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0.$$



LDA Pro/Con

Pros

- Multi-class version
- Simple to understand
- Usually doesn't overfit
- Works with much less data
- Very fast classification

Cons

- Simplistic Decision Boundaries
- Effected by points far from decision boundary