# *Machine Learning for Sensory Signals*

## Dimensionality Reduction

09-02-2017

# Principal Component Analysis

- First $M$ eigenvectors of data covariance matrix

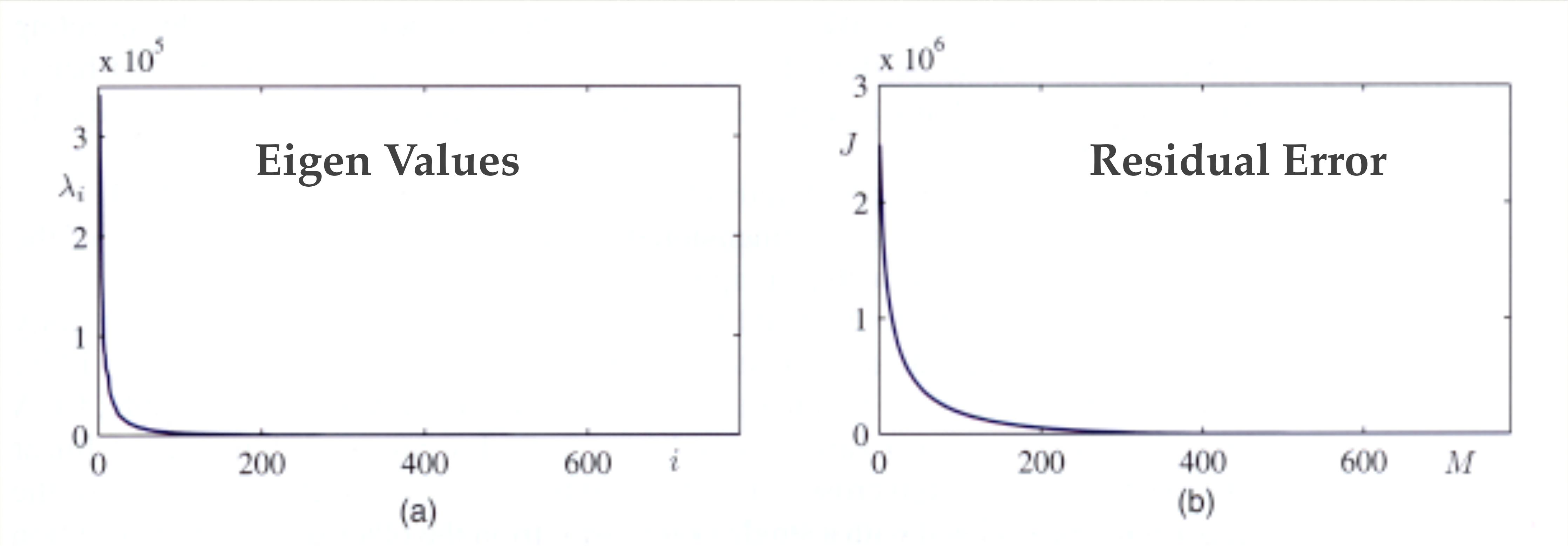
$$S = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

- Residual error from PCA

$$J = \sum_{i=M+1}^{D} \lambda_i$$

*PRML - C. Bishop (Sec. 12.1)*

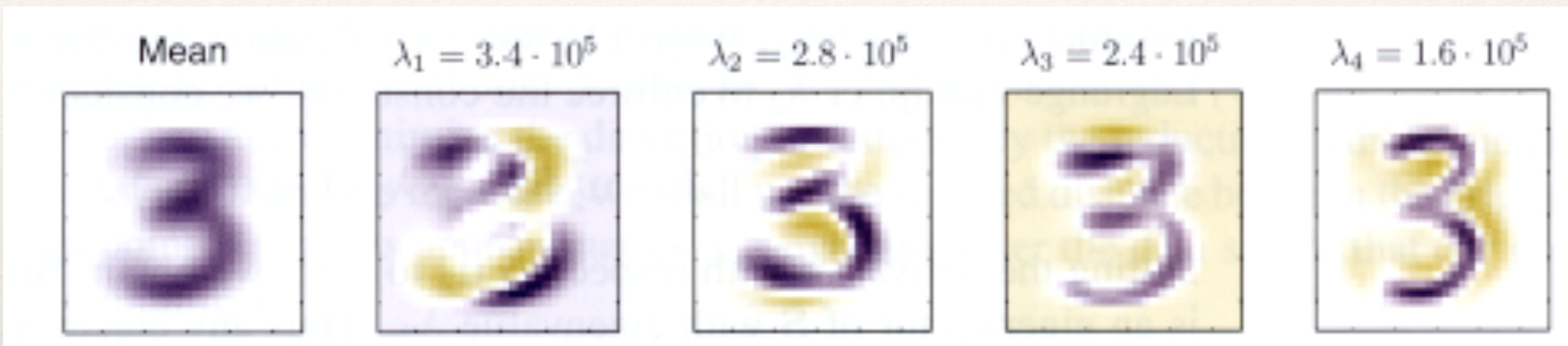
# PCA



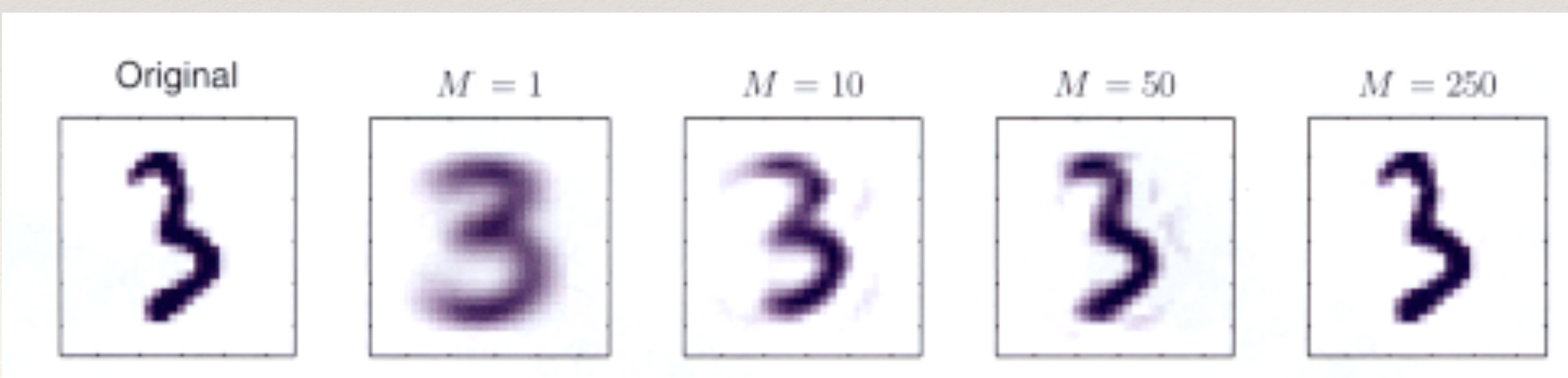

*PRML - C. Bishop (Sec. 12.1)*

# PCA - Reconstruction

**Eigenvectors**



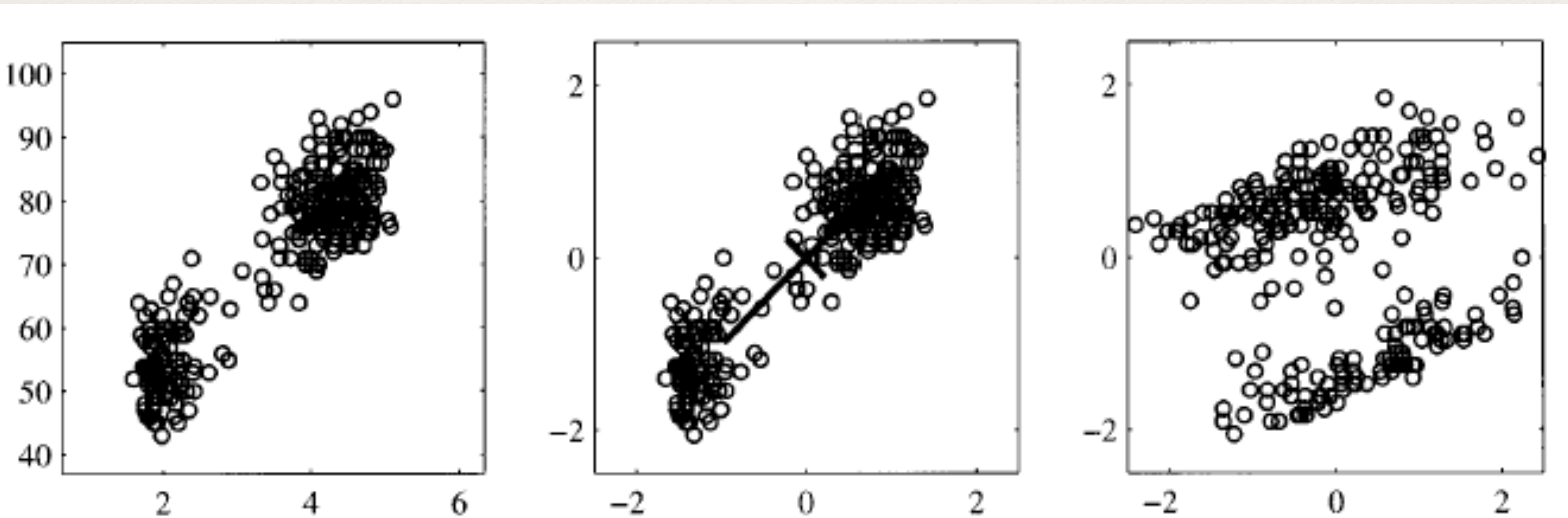

**PCA - Reconstruction**




*PRML - C. Bishop (Sec. 12.1)*

# Whitening the Data

**Original Data** **Mean Removed data** **Whitened data**




*PRML - C. Bishop (Sec. 12.1)*

# Linear Discriminant Analysis

Find a linear transform $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ with a criterion which maximizes the class separation

- Maximize the between class distance in the projected space while minimizing the within class covariance

$$J = \frac{\mathbf{w}^T \boldsymbol{S}_b \mathbf{w}}{\mathbf{w}^T \boldsymbol{S}_w \mathbf{w}}$$

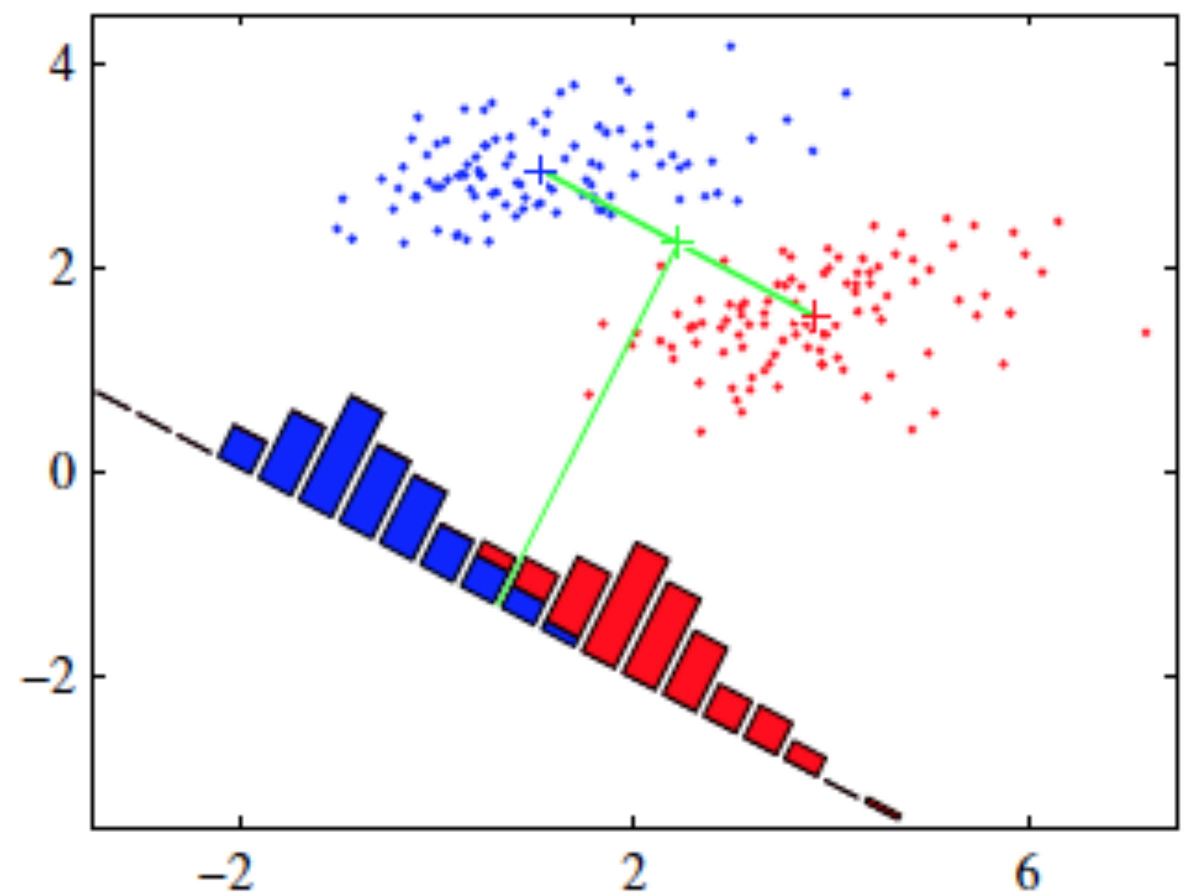
$$\boldsymbol{S}_b = \sum_{k=1}^{K} N_k(\mathbf{m}_k - \boldsymbol{m})(\mathbf{m}_k - \boldsymbol{m})^T \quad \boldsymbol{S}_w = \sum_{k=1}^{K} \sum_{n \in C_k} (\mathbf{x}_n - \boldsymbol{m}_k)(\mathbf{x}_n - \boldsymbol{m}_k)^T$$

- Generalized Eigenvalue problem
- Eigenvectors of $\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$

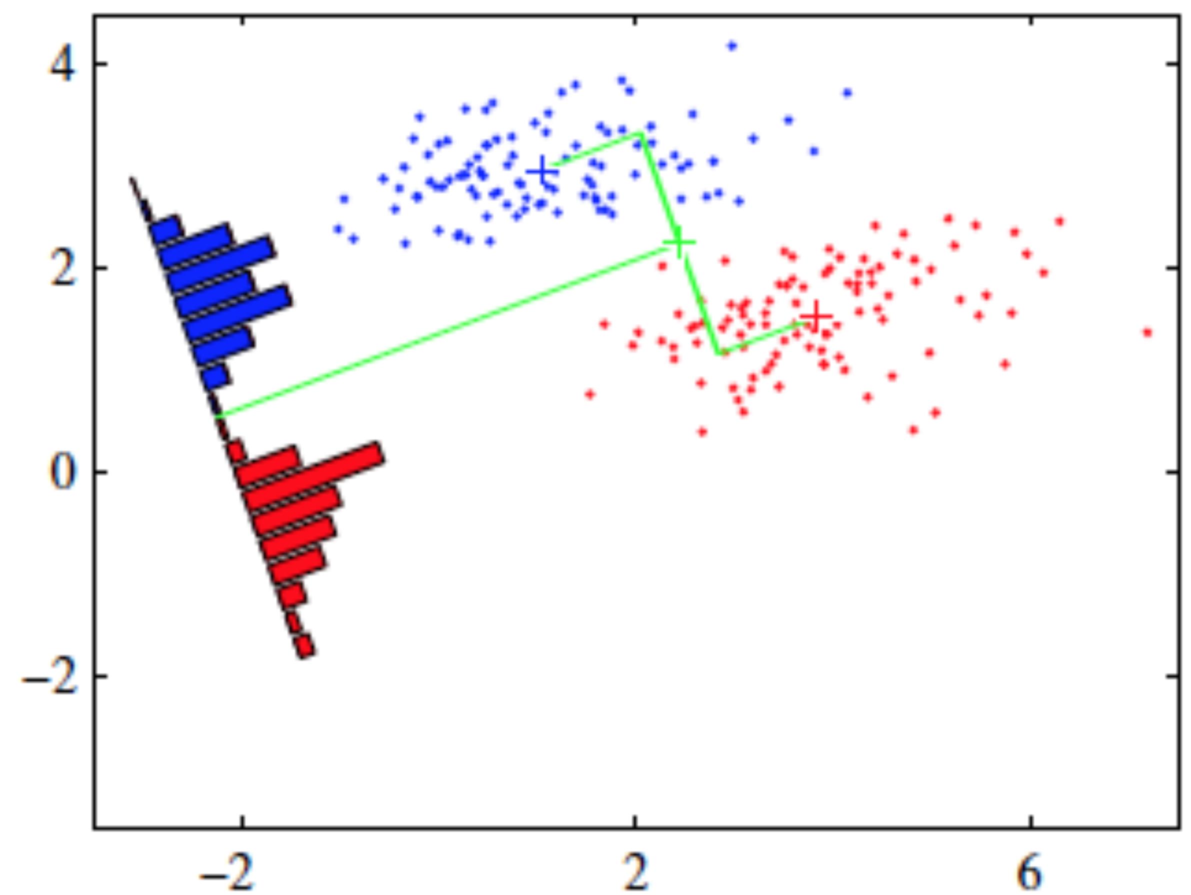
*PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)*

# Linear Discriminant Analysis

**Projecting on line joining means**

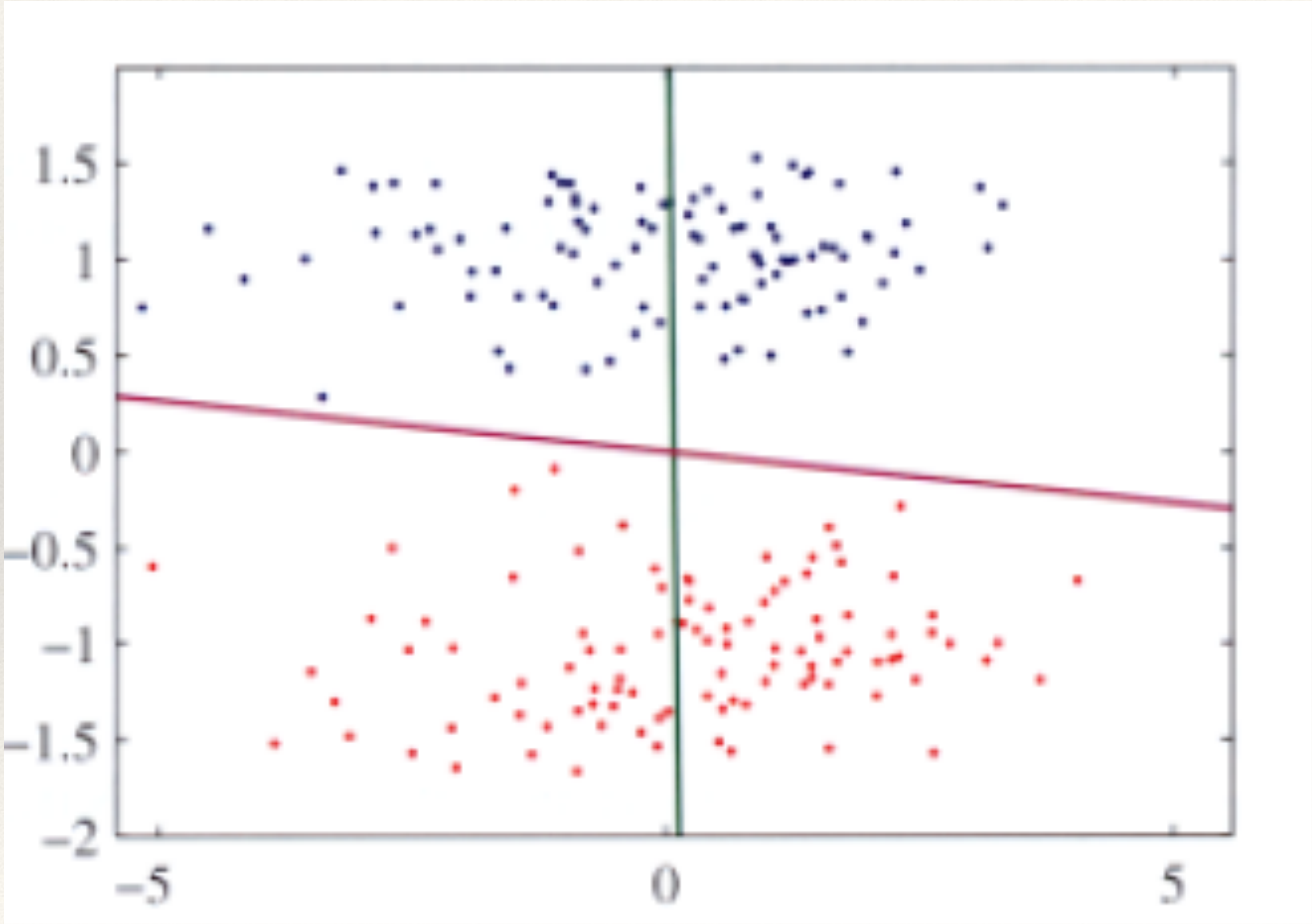



**Fisher Discriminant**




*PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)*

# PCA versus LDA

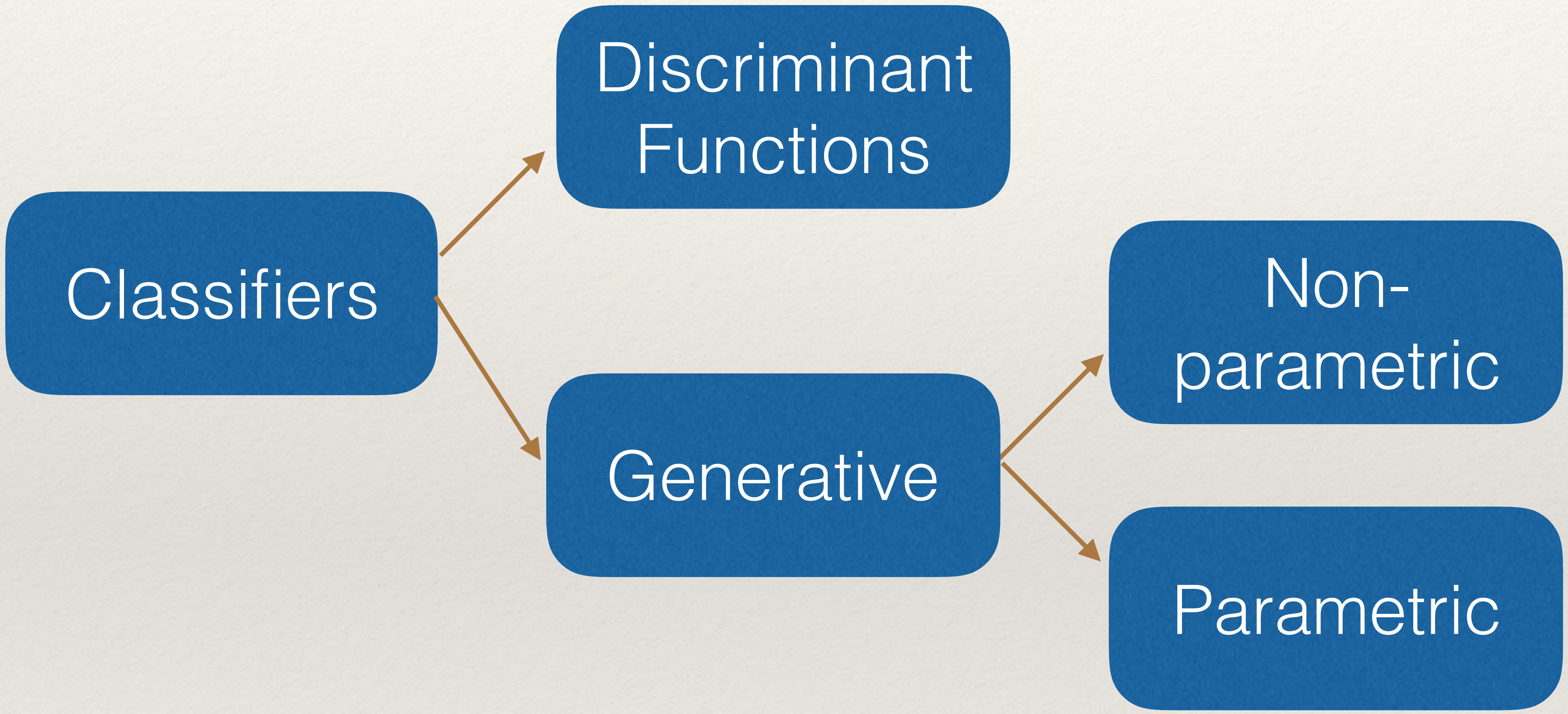



**PCA**

**LDA**

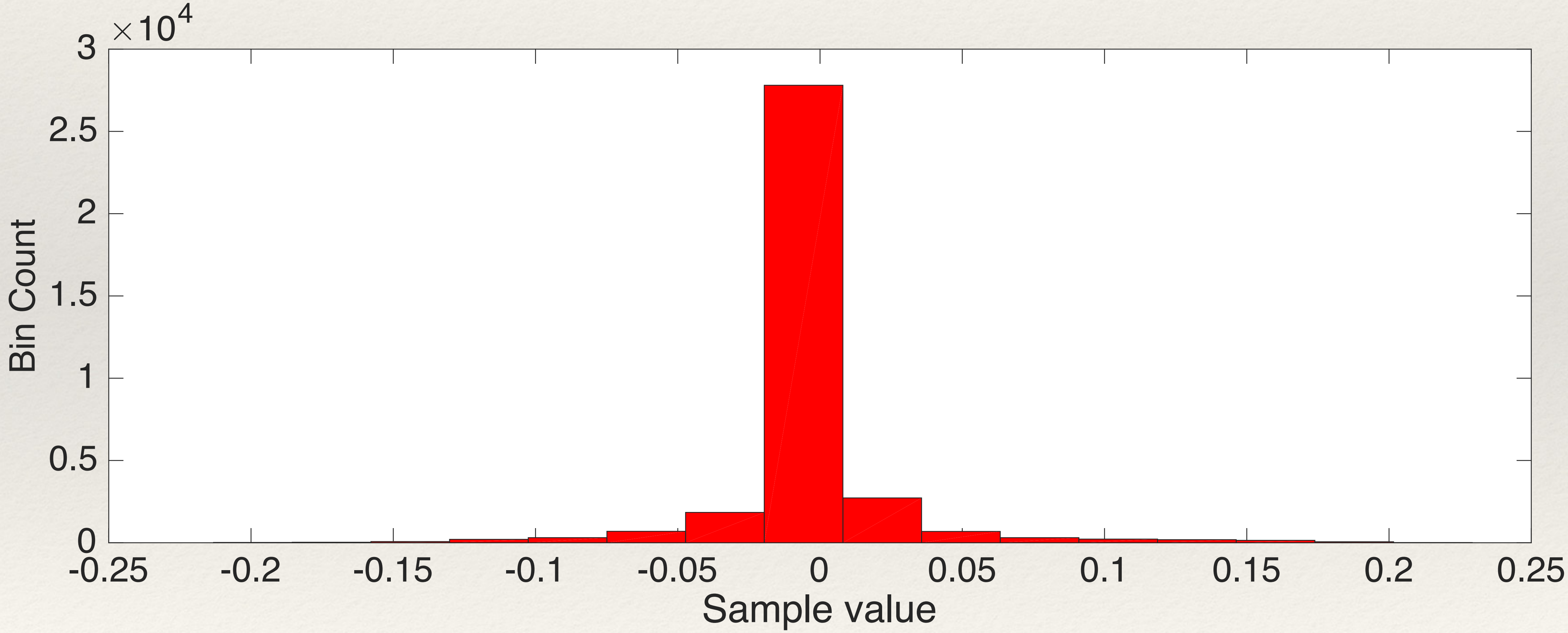
*PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)*

# Classifier Types



# Non-parametric Modeling

- Non-parametric models do not specify an apriori set of parameters to model the distribution. Example - Histogram

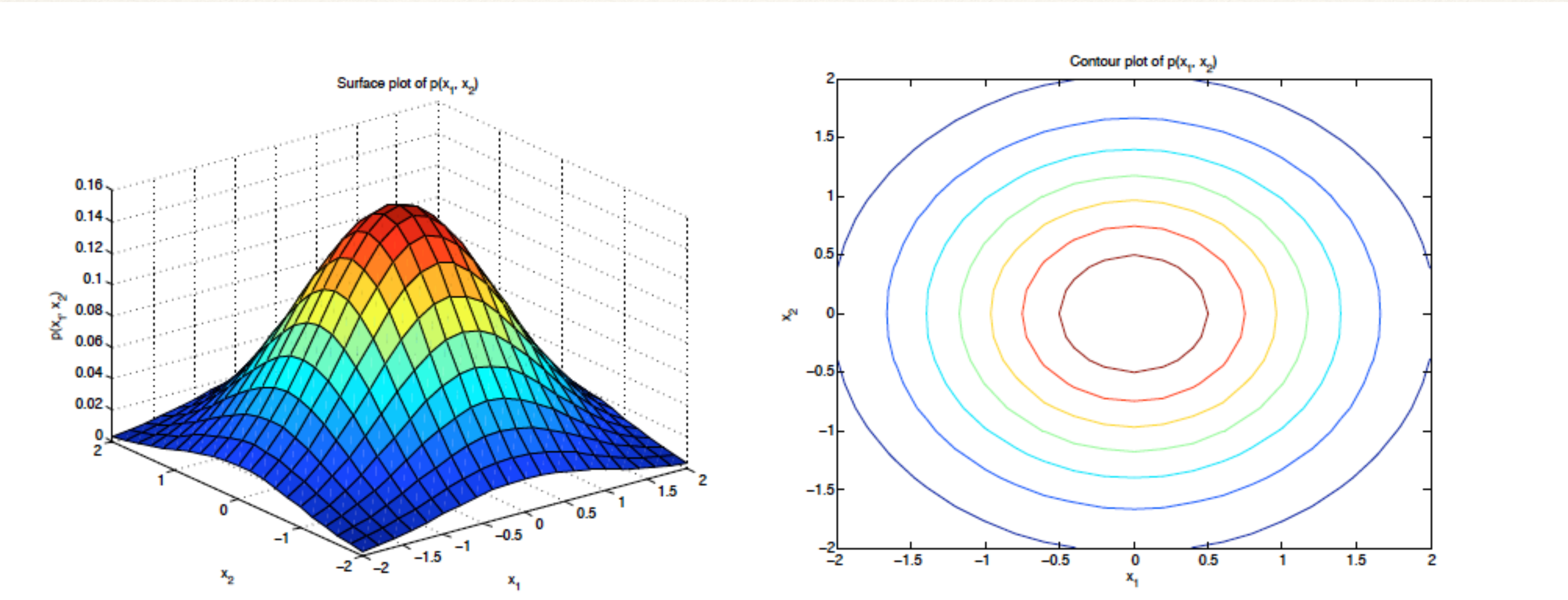



The density is not smooth and has block like shape.

# Non-parametric Modeling

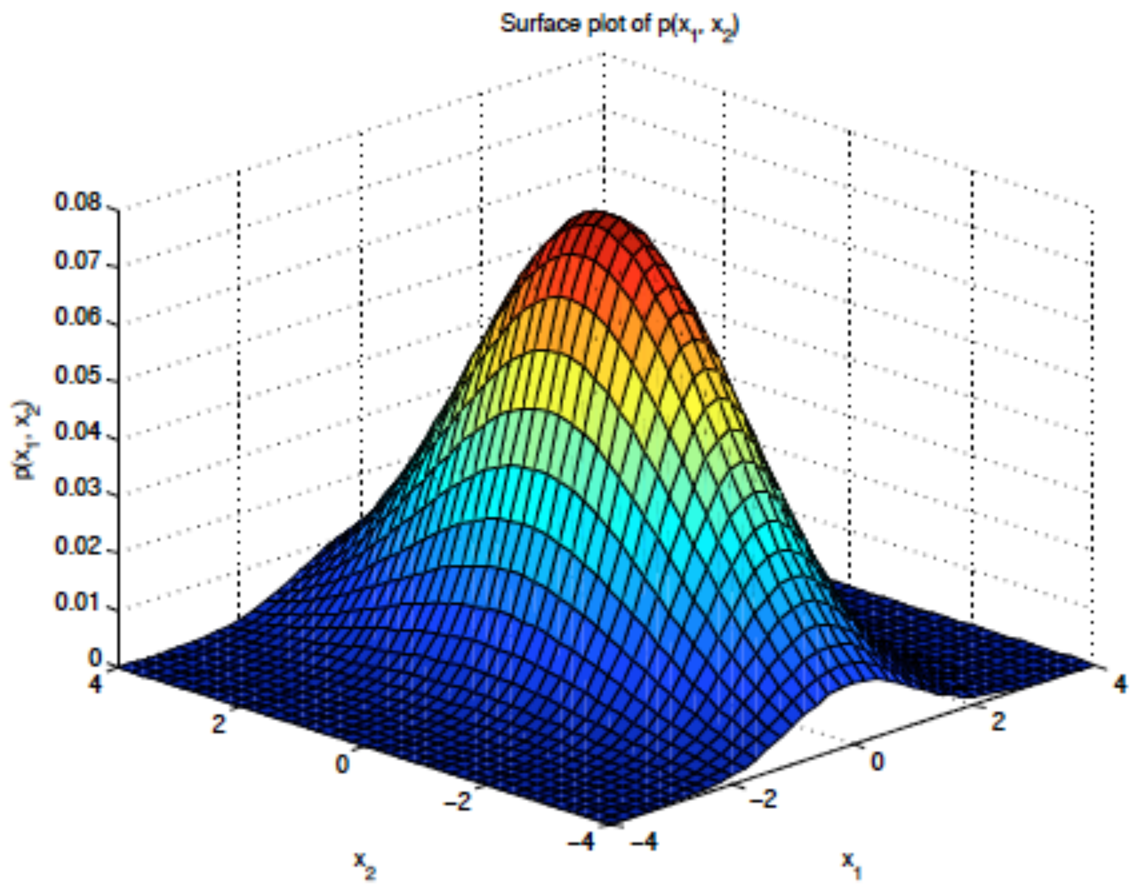
- Non-parametric methods are dependent on number of data points
    - Estimation is difficult for large datasets.
- Likelihood computation and model comparisons are hard.
- Limited use in classifiers

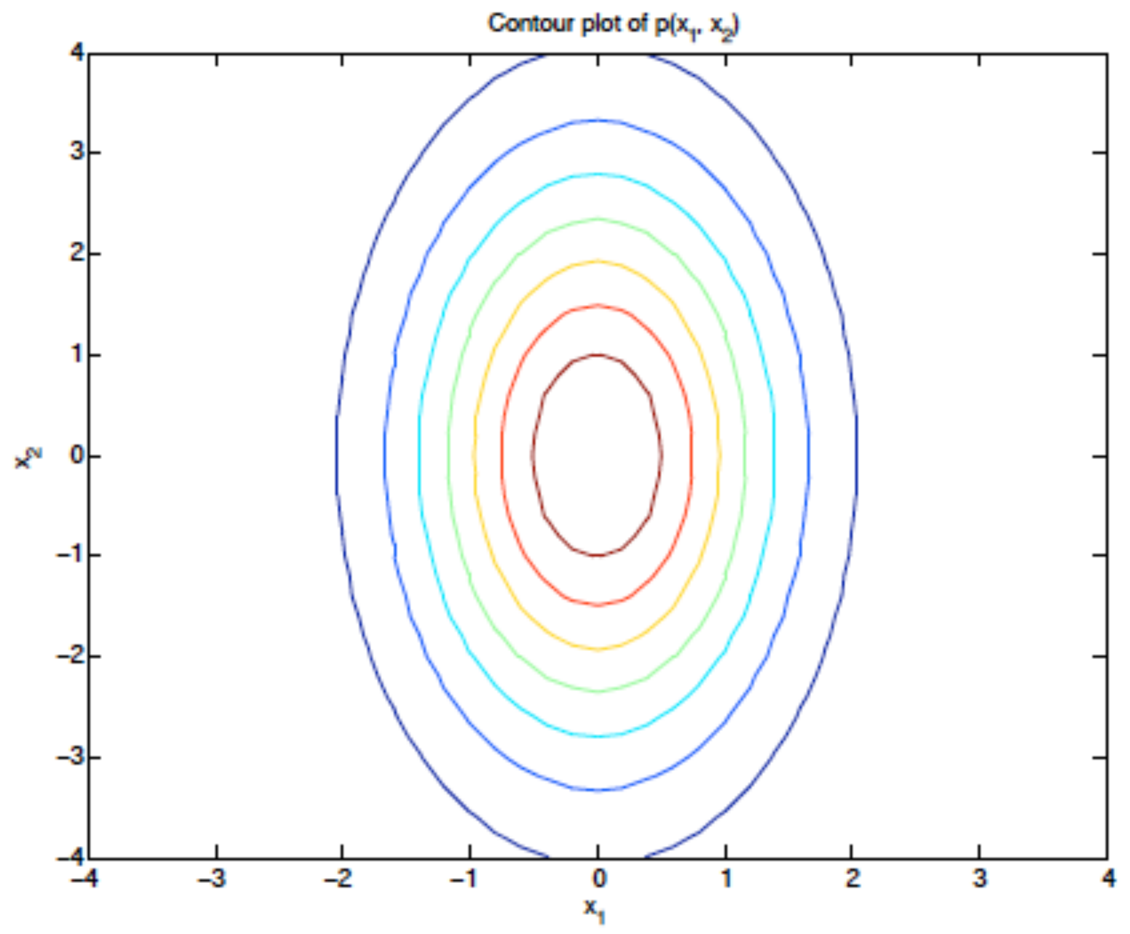
# Gaussian Distribution




Points of equal probability lie on on contour
Diagonal Gaussian with Identical Variance

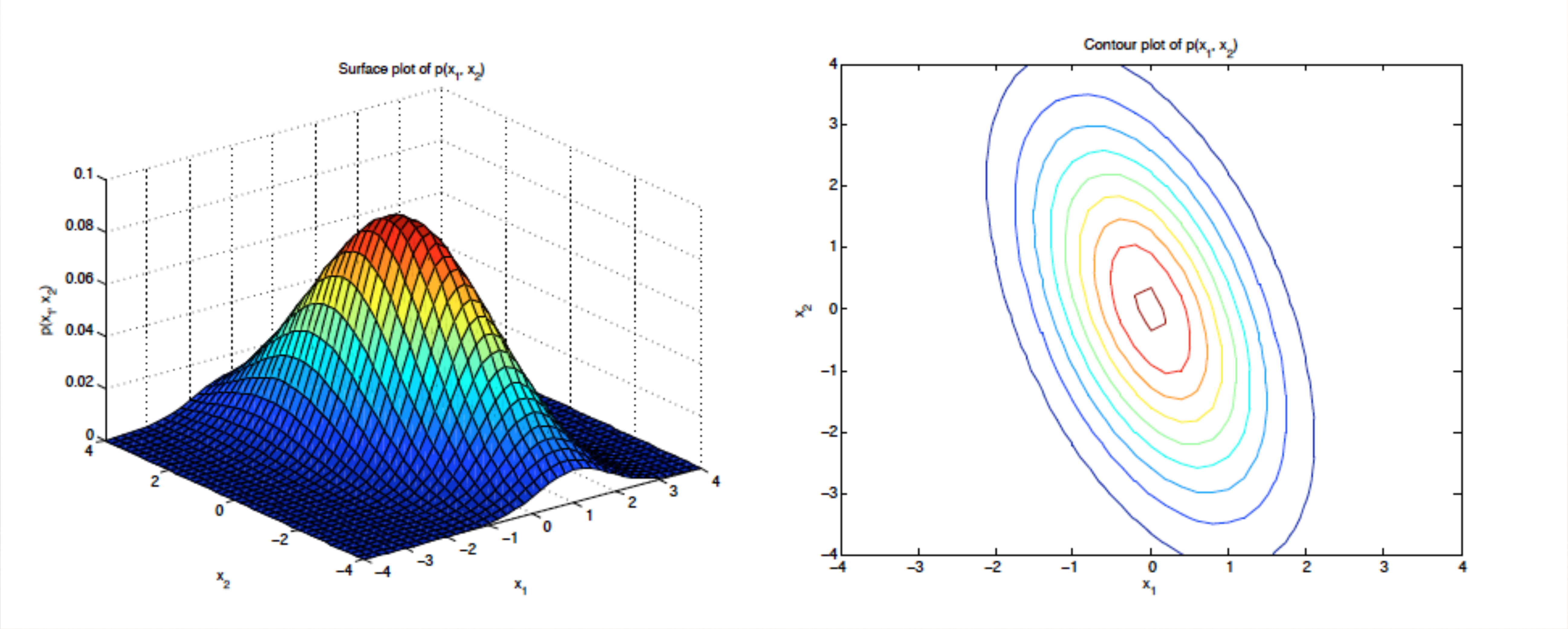
# Gaussian Distribution




Diagonal Gaussian with different variance

# Gaussian Distribution




Full covariance Gaussian distribution

# Gaussian Distribution