Module 28:
Performing MVPA

Performing MVPA

- The process of performing MVPA follows a series of steps:
  - Defining features and classes
  - Feature selection
  - Choosing a classifier
  - Training and testing the classifer
  - Examining results

Defining Features

- There are many possible choices of what information should be used as features.
  - Raw fMRI data over both space and time
  - Averaged fMRI data over a block
  - Beta values from a GLM analysis
  - Average of several voxels in an ROI

Defining Classes

- The choice of which class labels to use depends upon the research question.
  - Stimulus class
  - Subject response or decision
  - Any measurement that can be tied to an observation

Feature Selection

- In fMRI the number of features is typically many times larger than the number of observations.
- Hence, it is usually beneficial to reduce the number of features through feature selection.
- This could involve using only voxels from a particular ROI, dimension reduction techniques (SVD or PCA) or meta-analysis data.

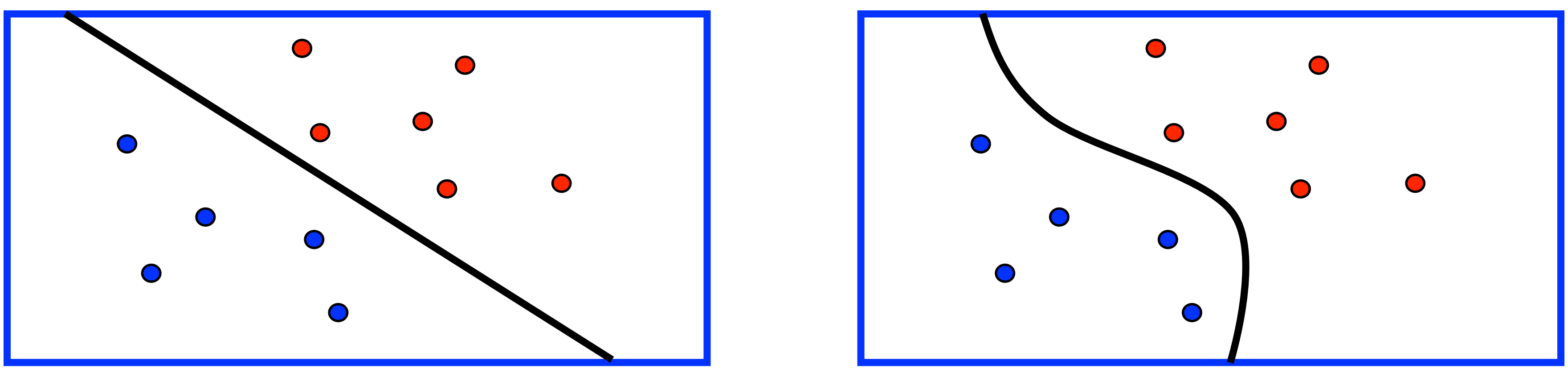
Feature Selection

- Note that it is not permissible to select voxels that appear to distinguish between classes using information from the entire data set.

- Information in the test data set may affect the learning of the classifier and bias subsequent accuracy measures.

Classifiers

- There are many types of classifiers to choose between that vary in the kinds of statistical relationships they can detect.
  – We often discriminate between linear and non-linear classifiers.

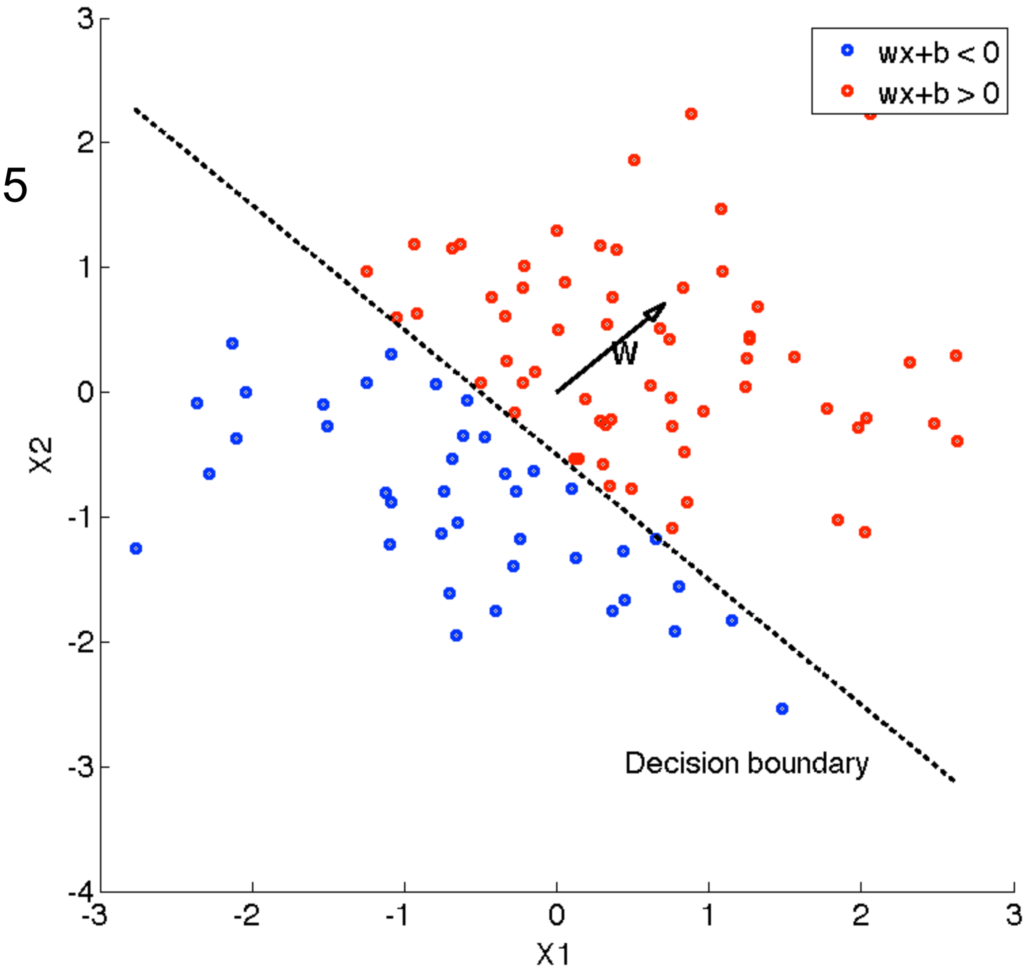


Linear Classifiers

- Linear classifier: $\underline{w}^T \underline{x} + b > 0$
  - In V-dimensions this defines a V-1 dimensional hyperplane
  - $\underline{w}$ is a V-dimensional vector of weights
  - b is a threshold
  - The inner product is zero when vectors are orthogonal, so the equation $\underline{w}^T\underline{x} = 0$ defines a line orthogonal to $\underline{w}$.

Example

$w = [1\ 1]^T$, $b = .5$



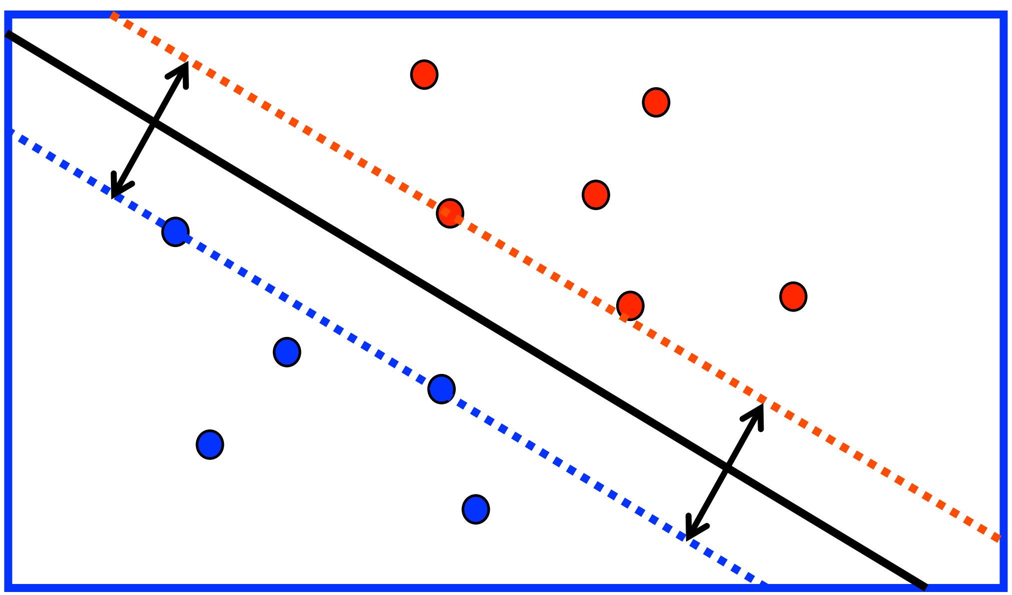
Linear Classifiers

- There exist many types of linear classifiers.

- Some examples include:
  - Linear Support Vector Machines (SVM)
  - Logistic Regression (LR)
  - Gaussian Naive Bayes (GNB)
  - Fisher's Linear Discriminant Analysis (LDA)

SVM

- SVMs maximize the margin around the separating hyperplane.
  - If there are no points near the decision surface, then there are no uncertain classification decisions



Support Vector Machines

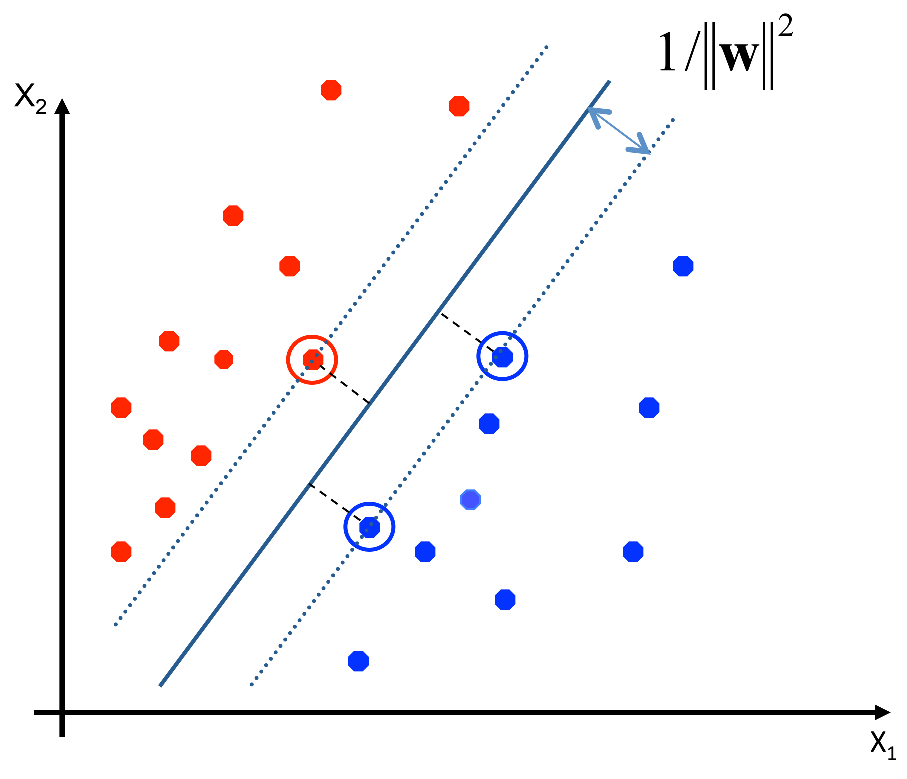
- Consider (x_i, y_i), $i=1,\ldots N$, where $x_i \in R^p$ and $y_i \in \{-1,1\}$.

- Solve the convex optimization problem:

Minimize: $$\frac{1}{2}\|\mathbf{w}\|^2$$

Subject to $$y_i(\mathbf{w}^T\mathbf{x}_i - b) \leq 1$$

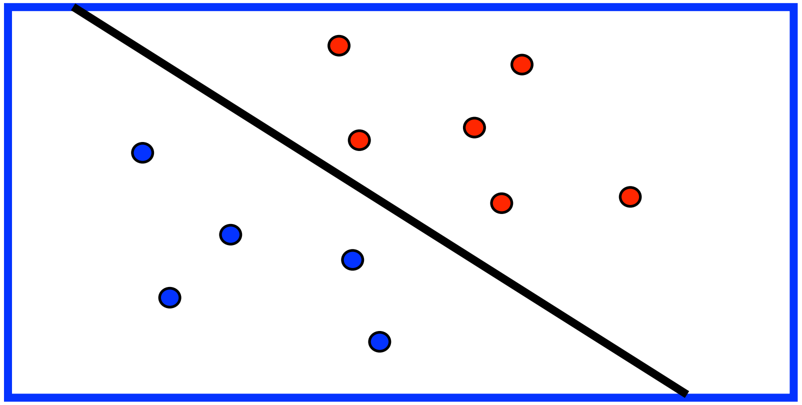
Solving SVMs is a quadratic programming problem



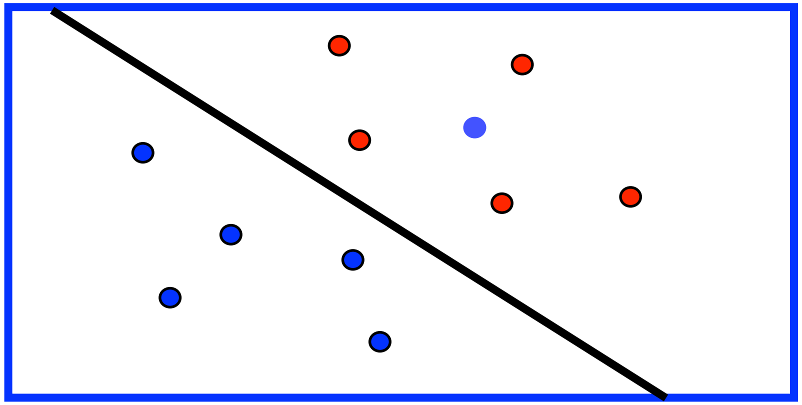
Separable

- Data sets that can be separated exactly by a linear boundary are said to be linearly separable.

Linearly separable



Not linearly separable



Slack Variable

- When the data is not linearly separable, we may still use a linear classifier by allowing certain data points to be on the 'wrong side' of the boundary.
  - However, they incur a penalty that increases with their distance from the boundary.

- To implement this we can introduce slack variables $\xi_i \geq 0$ to allow misclassification of difficult or noisy observations.

Support Vector Machines

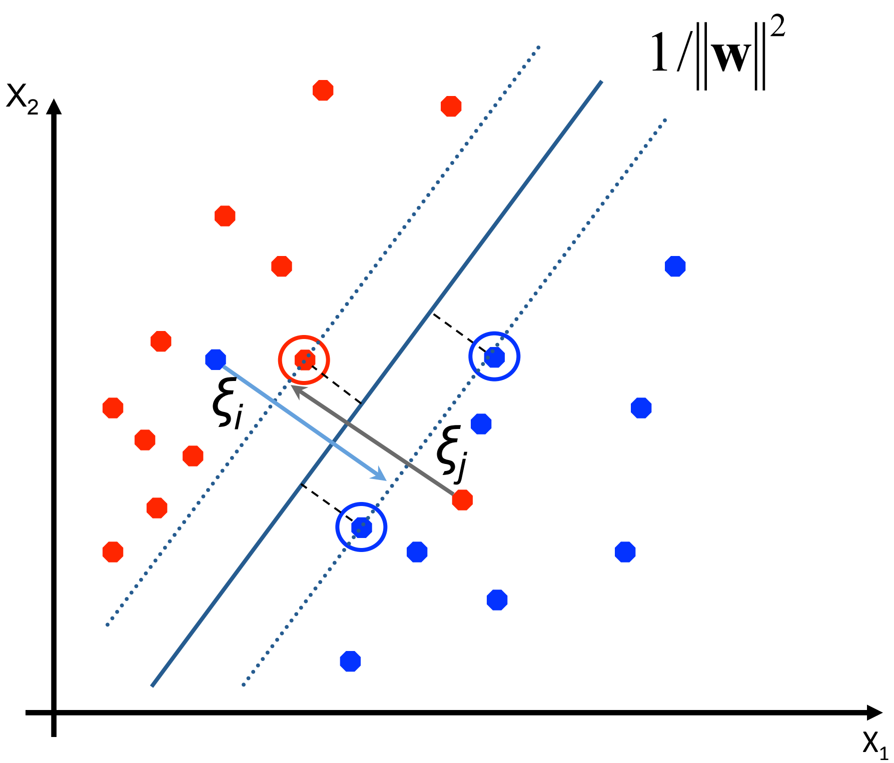
- Consider (x_i, y_i), $i=1,\ldots N$, where $x_i \in R^p$ and $y_i \in \{-1,1\}$.

- Solve the convex optimization problem:

Minimze: $$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

Subject to $$y_i(\mathbf{w}^T\mathbf{x}_i - b) \geq 1 - \xi_i$$

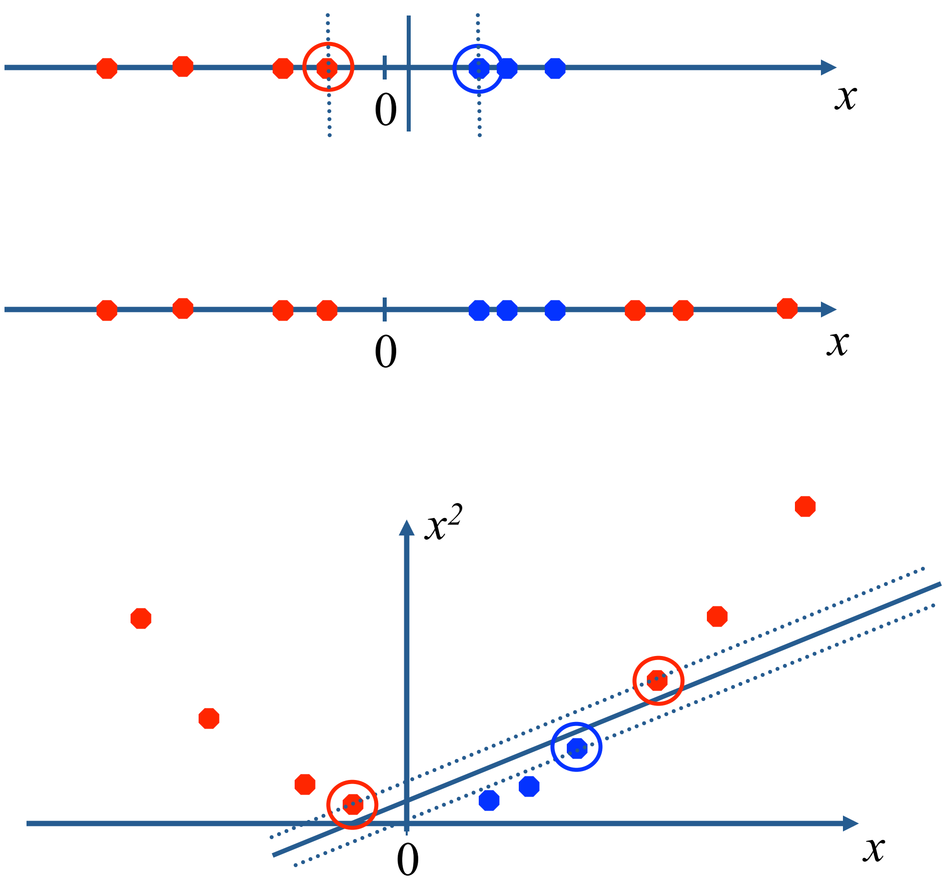
$$\xi_i \geq 0$$



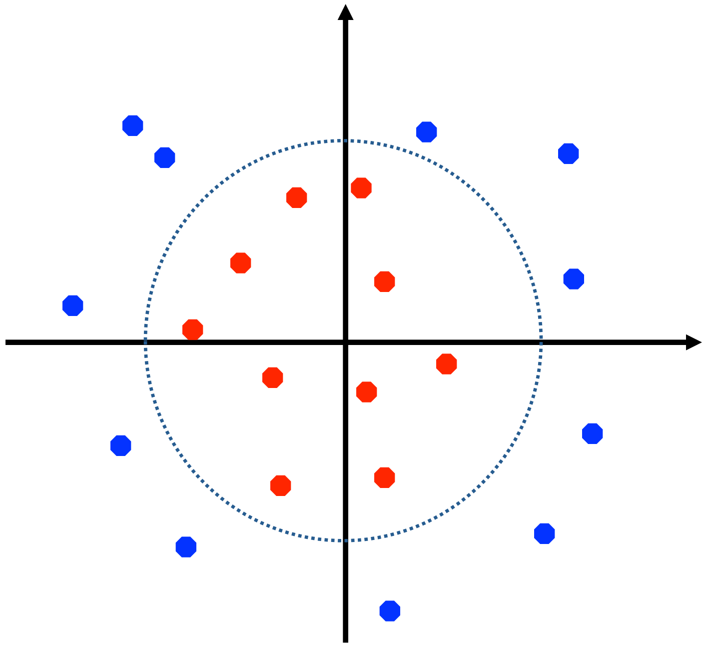
Incorporating Nonlinearity

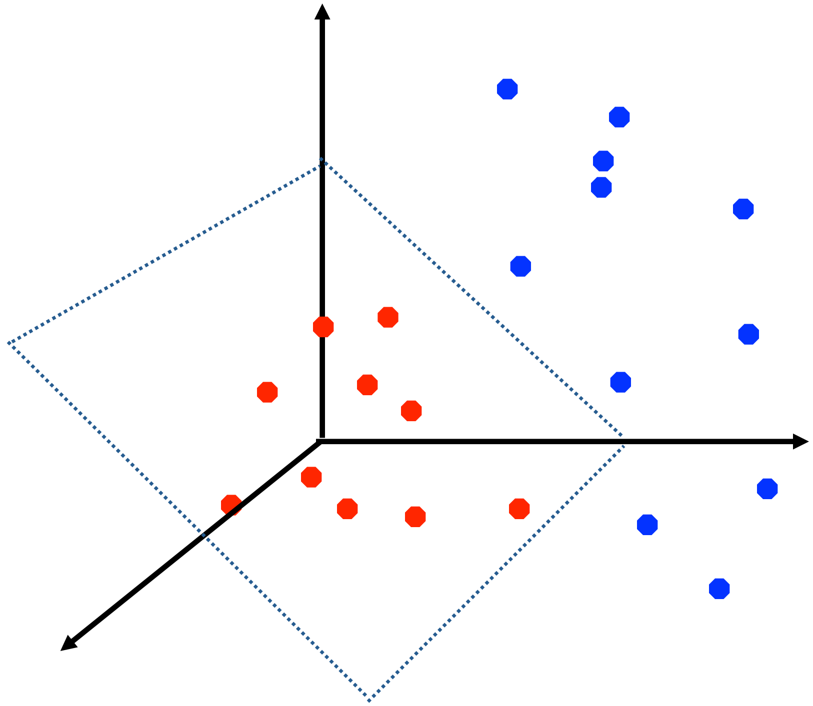
- Data sets that are linearly separable tend to be easy to work with.
- If the data isn't linearly separable we can always introduce slack variables.
- Another option is to map the data to a higher-dimensional space where the training set is separable with a linear classifier.

Nonlinear Classification



Nonlinear Classification





End of Module

@fMRIstats