Support Vector Machines and Speaker Verification

David Cinciruk

March 6, 2013

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

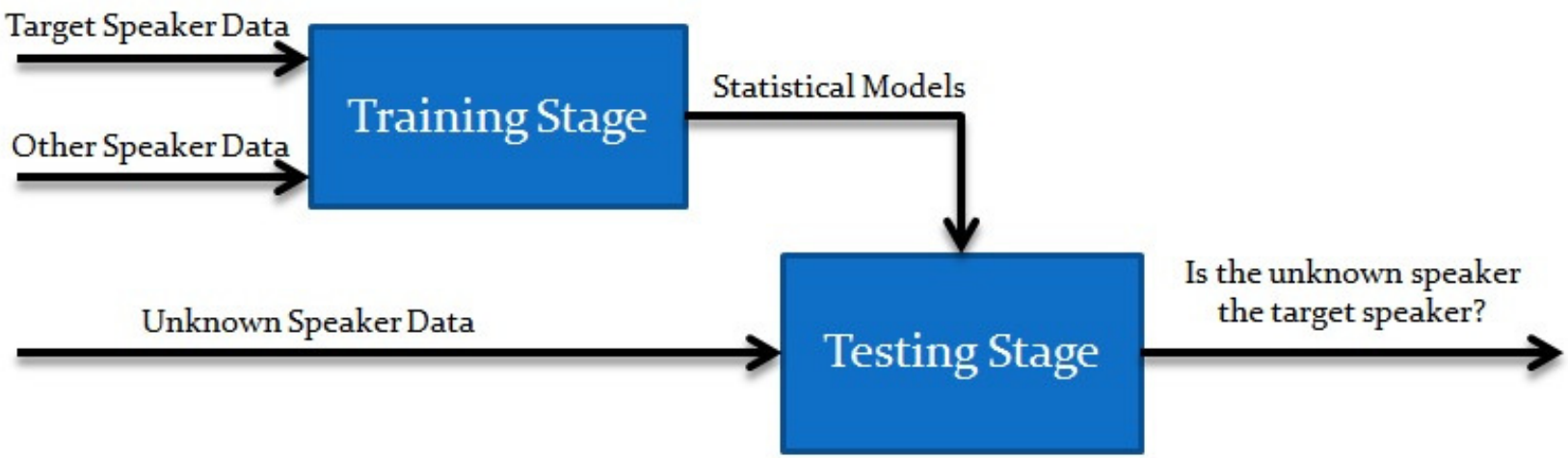
Derivation of SVM Equations

Soft Margin

Nonlinear Classification

SVMs in Speaker Verification
- Examples of Kernels Used in Speaker Verification
- How to Perform

Review of Speaker Verification



- Speaker Verification as discussed before was done using Gaussian Mixture Models
- One popular way to perform speaker verification is with Support Vector Machines

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

Soft Margin

Nonlinear Classification

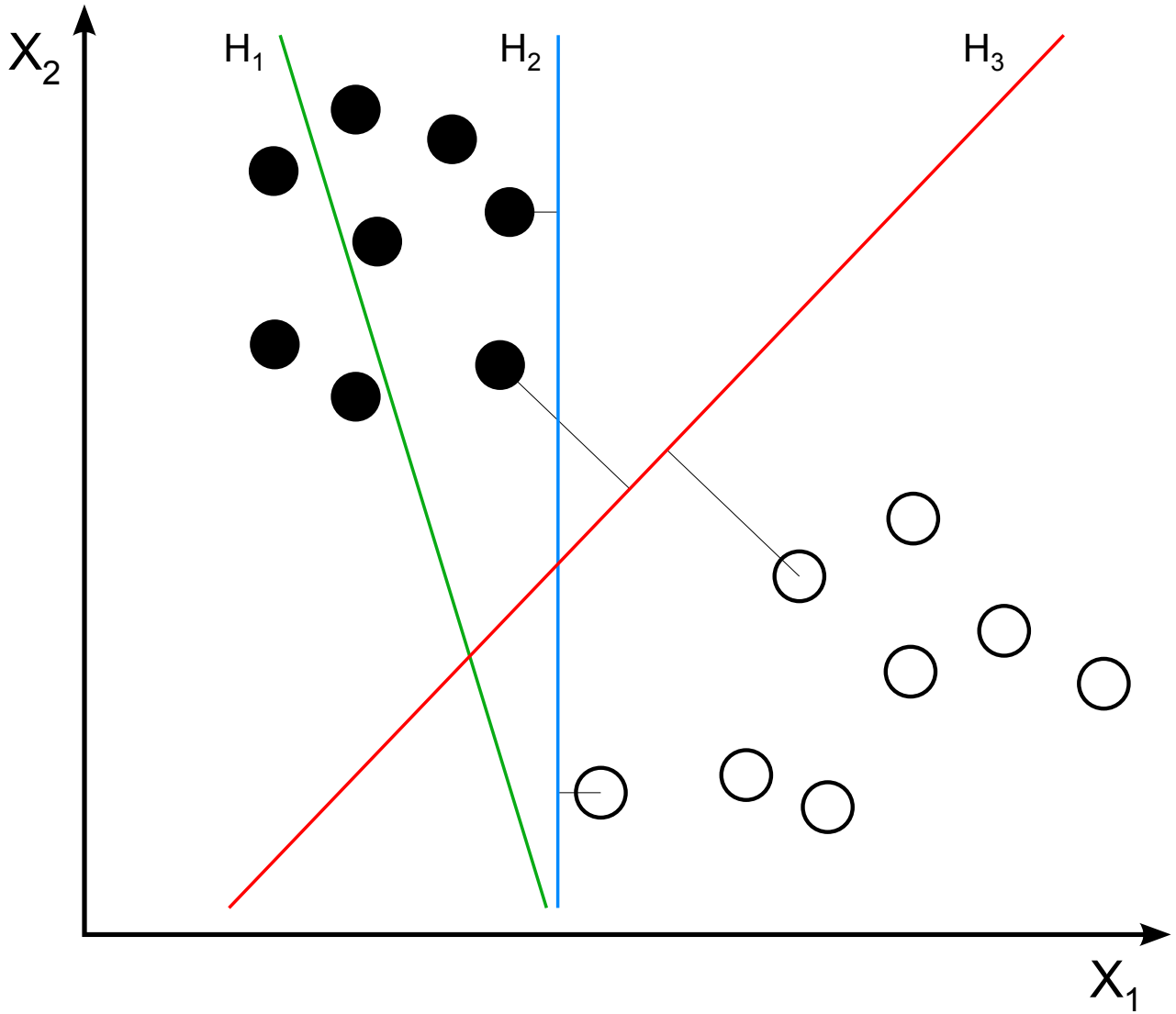
SVMs in Speaker Verification

- Examples of Kernels Used in Speaker Verification
- How to Perform

Motivation

- Training Data from two different classes are linearly separable.
- Want to classify unknown testing data into one of the classes
- Since training data is linearly separable, can create a line that separates the data
- Need to be able to classify testing data with minimum errors

A Pictoral Overview of Support Vector Machines



Kernel Functions

- Requires kernel trick to work with non-linearly separable training data
- Kernel Trick - Mapping items from a set S into an inner product space V without having to compute the mapping
- Data can be projected to higher dimensions using kernels
- In the linear case, kernel is:

$$k(x_i, x_j) = x_i \cdot x_j \tag{1}$$

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

Soft Margin

Nonlinear Classification

SVMs in Speaker Verification
- Examples of Kernels Used in Speaker Verification
- How to Perform

Formulation of the Problem

- Consider a set of data points of the form

$$\mathcal{S} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\}\} \tag{2}$$

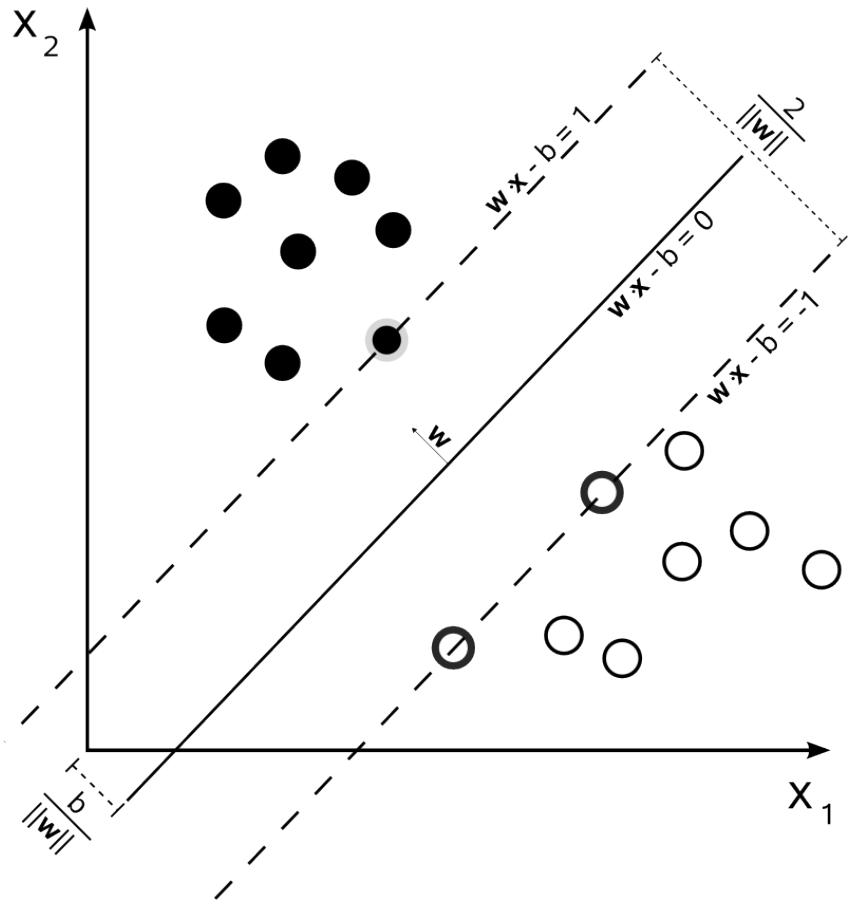
- y_i - indicator of which class a point $\mathbf{x}_i$ is of
- Need to a find a hyperplane that maximally separates data. Hyperplane of form:

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \tag{3}$$

- Parameter $\frac{b}{||\mathbf{w}||}$ denotes affine offset from origin along normal vector $\mathbf{w}$

Selecting the Seperating Hyperplane

- First case: Linearly Separable Training Data
- Select two hyperplanes as shown
- Region between: "the margin"
- Hyperplanes are of form on picture
- Distance between the two is $\frac{2}{||\mathbf{w}||}$
- Thus, need to minimize $||\mathbf{w}||$



Formulating the Optimization Problem

- Want to prevent points from falling into the margin. Thus add constraints
- For $\mathbf{x}_i$ of the first class

$$\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \tag{4}$$

- For $\mathbf{x}_i$ of the second class

$$\mathbf{w} \cdot \mathbf{x}_i - b \leq -1 \tag{5}$$

- Or can be written together as

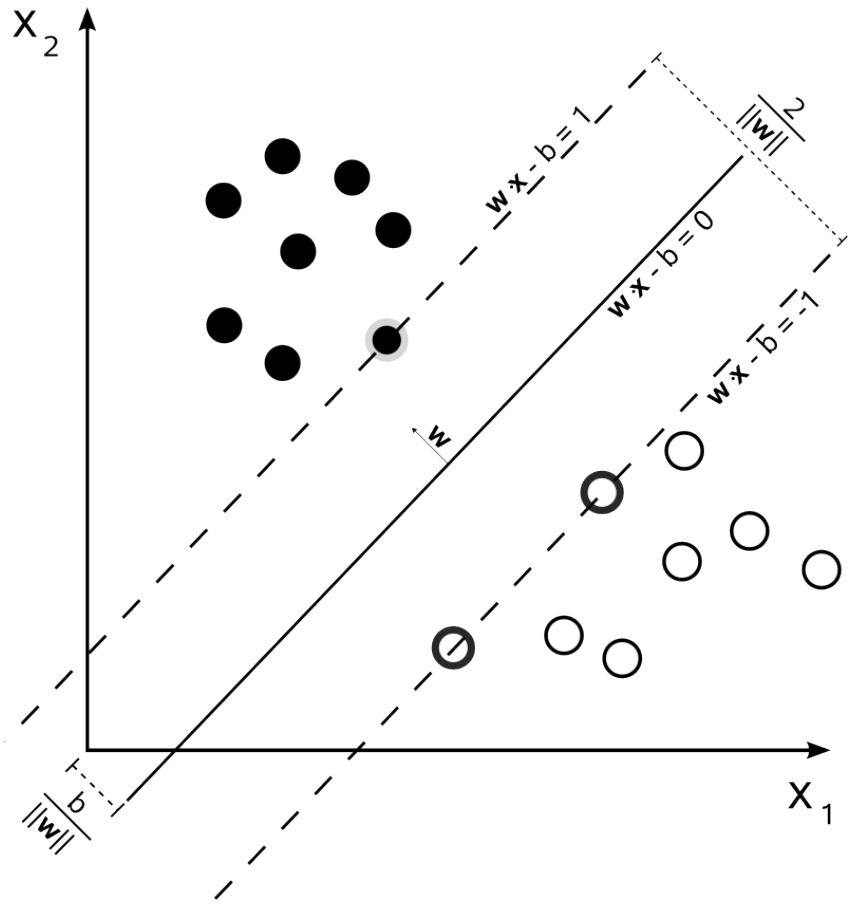
$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 \tag{6}$$

The Optimization Problem

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 \qquad (7)$$

subject to

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 \qquad (8)$$

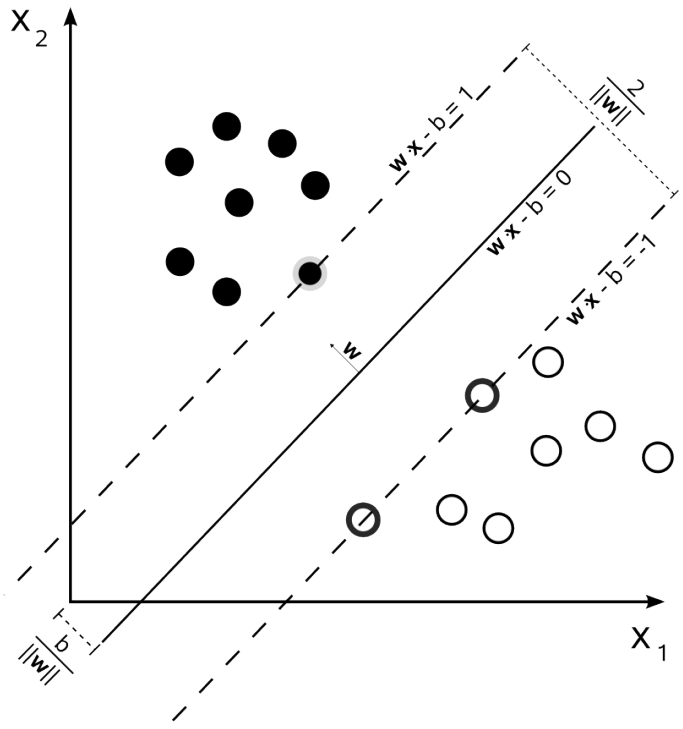


The Primal Problem

- The Lagrangian can be defined as

$$L_P := \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i - b) - 1] \tag{9}$$

- Primal Problem is convex since objective is convex and constraints are linear.



The Primal Problem

Solving the gradient gives us

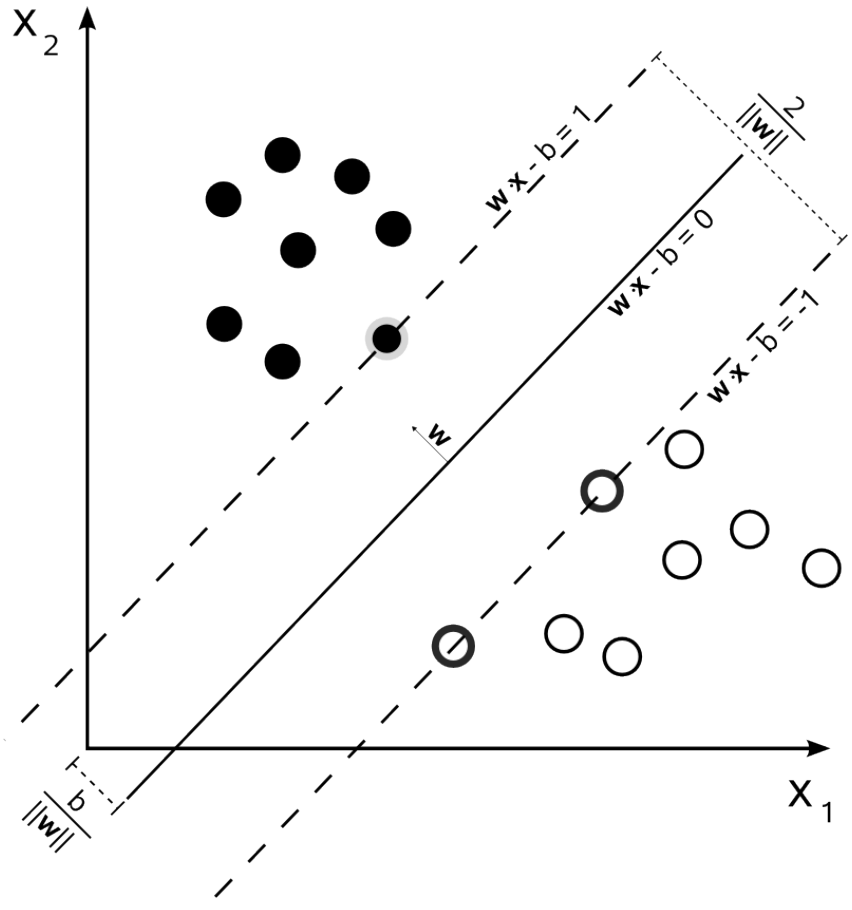
$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \qquad (10)$$

$$\sum_i \alpha_i y_i = 0 \qquad (11)$$

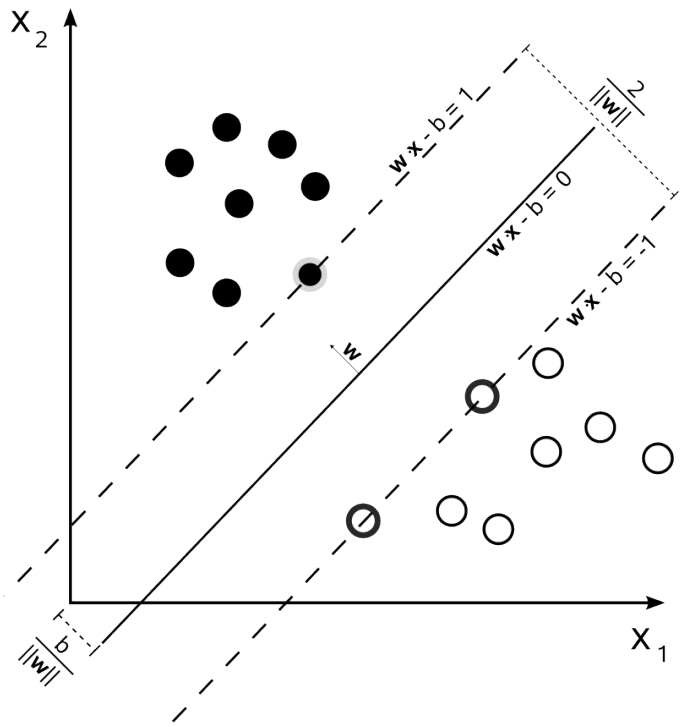
and

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (\mathbf{w} \cdot \mathbf{x}_i - y_i) \qquad (12)$$

where i is taken over the support vectors (the vectors on the edge of the boundary)



The Dual Problem



- Since primal problem is convex, there is no duality gap
- Dual Lagrangian given as

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{13}$$

KKT Conditions

$$\frac{\partial}{\partial w_v} L_P = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad v = 1, \ldots, d \tag{14}$$

$$\frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \tag{15}$$

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad i = 1, \ldots, l \tag{16}$$

$$\alpha_i \geq 0 \quad \forall i \tag{17}$$

$$\alpha_i(y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0 \quad \forall i \tag{18}$$

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

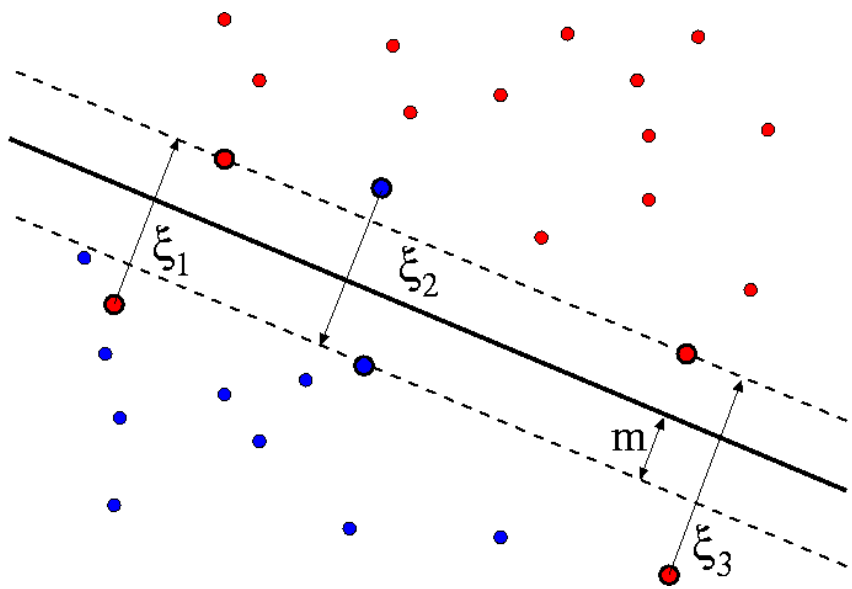
Soft Margin

Nonlinear Classification

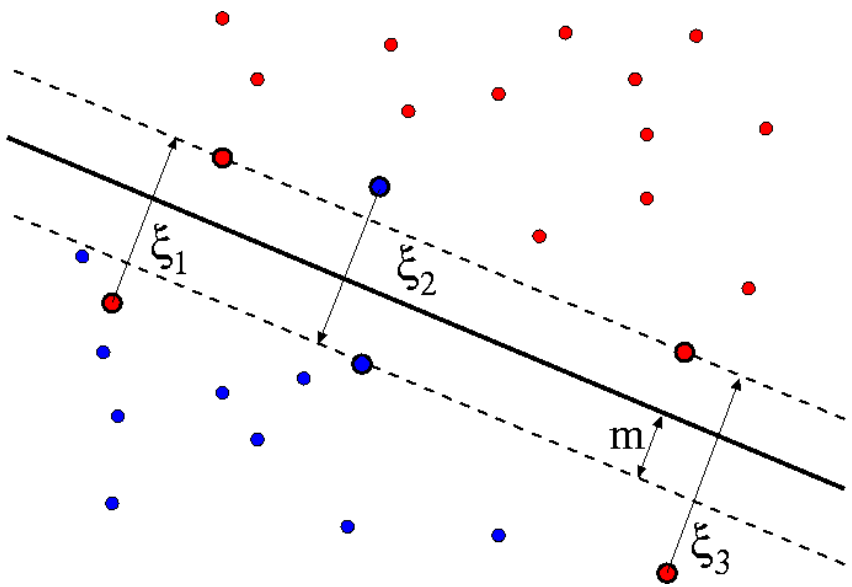
SVMs in Speaker Verification
- Examples of Kernels Used in Speaker Verification
- How to Perform

Unseparable Data

- Sometimes data from one class is mixed with data from other
- Can project with kernels but may overfit
- Want to allow for errors



Optimization Problem for Soft Margin



- Introduce linear penalty function featuring slack variable ξ_i which measure the degree of misclassification of the data
- Optimization problem becomes

$$\min_{\mathbf{w},\xi,b}\{\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i\} \quad (19)$$

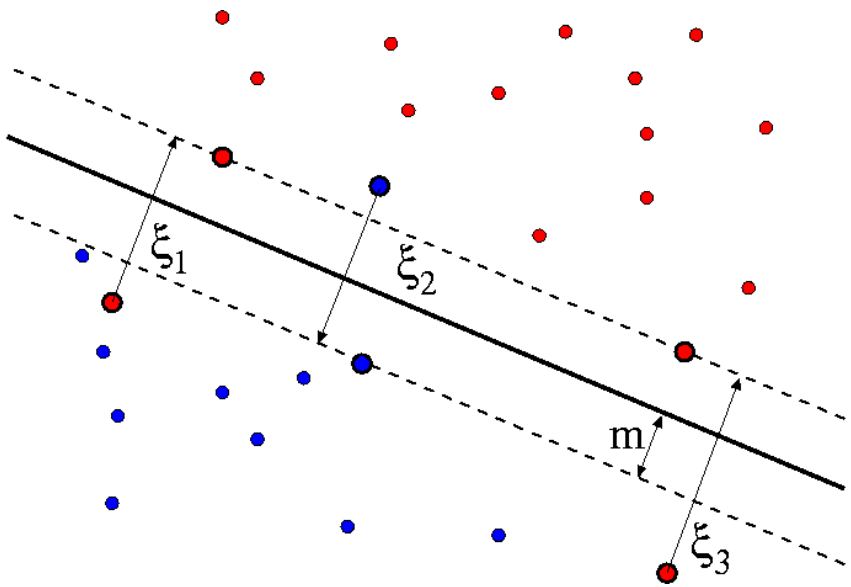
subject to

$$y_i(\mathbf{w}\cdot\mathbf{x}_i - b) \geq 1 - \xi_i \quad \xi_i \geq 0 \quad (20)$$

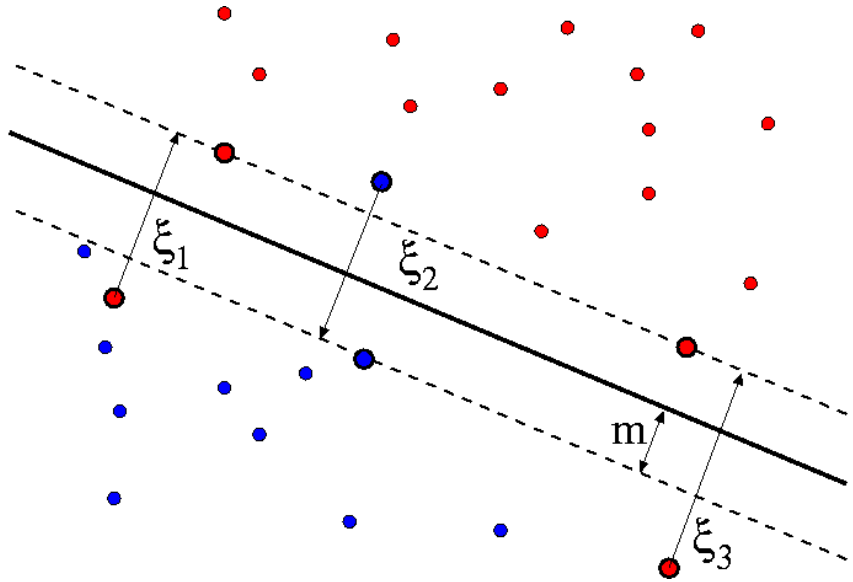
Primal Lagrangian

$$L_P = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}\cdot\mathbf{x}_i - b) - 1 + \xi_i] - \sum_{i=1}^{n}\beta_i\xi_i \quad (21)$$

with $\alpha_i, \beta_i \geq 0$



Dual Lagrangian



$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{22}$$

subject to

$$0 \leq \alpha_i \leq C \tag{23}$$

and

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{24}$$

KKT Conditions for Soft Margin

$$\frac{\partial L_P}{\partial w_v} = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \tag{25}$$

$$\frac{\partial L_P}{\partial b} = -\sum_i \alpha_i y_i = 0 \tag{26}$$

$$\frac{\partial L_P}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \tag{27}$$

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i \geq 0 \tag{28}$$

$$\xi_i \geq 0 \tag{29}$$

$$\alpha_i \geq 0 \tag{30}$$

$$\mu_i \geq 0 \tag{31}$$

$$\alpha_i[y_i(\mathbf{w} \cdot \mathbf{x}_i - b) - 1 + \xi_i] = 0 \tag{32}$$

$$\mu_i \xi_i = 0 \tag{33}$$

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

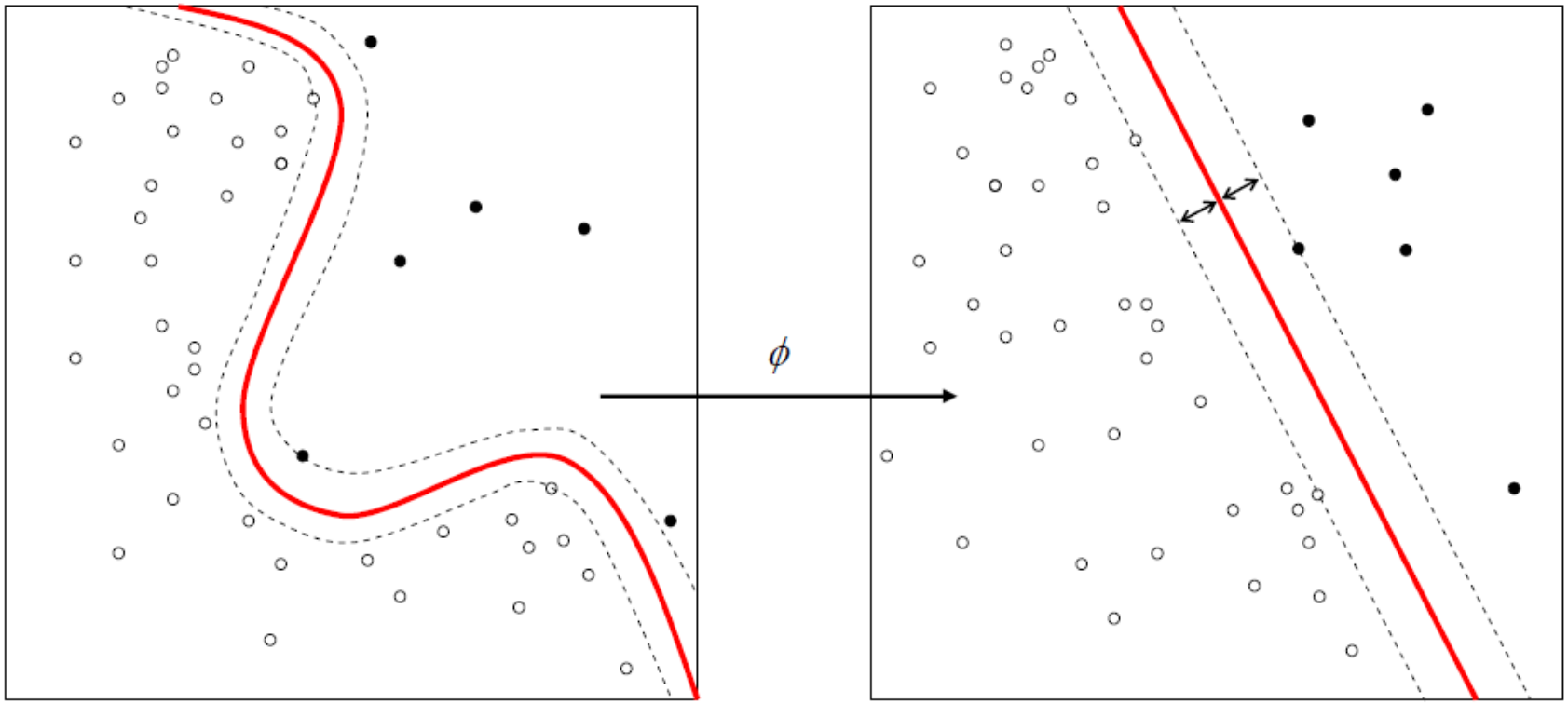
Derivation of SVM Equations

Soft Margin

Nonlinear Classification

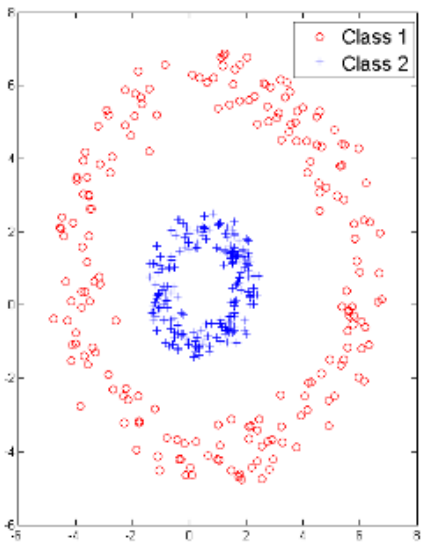
SVMs in Speaker Verification
- Examples of Kernels Used in Speaker Verification
- How to Perform

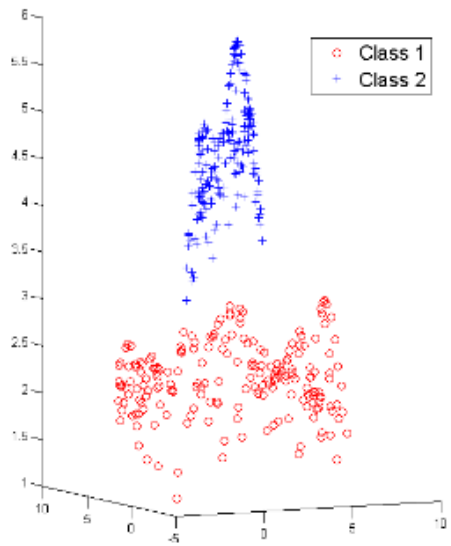
Motivation



- Use Kernels to Transform Nonlinearly separable data into linearly separable data
- Project Data to Higher Dimensions

Kernel Function





- Kernel function is dot product of two vectors projected into another space.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi((x)_i) \cdot \phi((x)_j) \tag{34}$$

- SVM equations only require dot product of two vectors. Can work with infinite dimensional projections

Calculating the Decision Boundary

- Can use same optimization as before, substituting $K(\mathbf{x}_i, \mathbf{x}_j)$ for $\mathbf{x}_i \cdot \mathbf{x}_j$
- New data points can be classified by following equation

$$f(x) = \sum_{i=1}^{N_S} \alpha_i y(i) K(\mathbf{s}_i, \mathbf{x}) + b \tag{35}$$

where $\mathbf{s}_i$ is the support vectors

Mercer's Condition

- Kernels must be positive semidefinite to obtain uniform convergence to a solution for all training sets
- For any $g(\mathbf{x})$ such that $|\int g(\mathbf{x})d\mathbf{x}| < \infty$

$$\int K(\mathbf{x}, \mathbf{y})g(\mathbf{x})g(\mathbf{y})d\mathbf{x}d\mathbf{y} \geq 0 \tag{36}$$

If Mercer Condition is Not Achieved

- Training data may cause Hessian to be indefinite
- No solution can be found generally
- May find solution for some set of training vectors

Examples of Kernels

- Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d \tag{37}$$

- Gaussian Radial Basis Function

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2} \tag{38}$$

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

Soft Margin

Nonlinear Classification

SVMs in Speaker Verification

- Examples of Kernels Used in Speaker Verification
- How to Perform

Typical Question

- What Kernel to Use?
- What Space to Model Data in?

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

Soft Margin

Nonlinear Classification

SVMs in Speaker Verification

- Examples of Kernels Used in Speaker Verification
- How to Perform

Radial Basis Functions

- Use standard radial basis function for kernel
- Run on MFCCs: each one a data point
- Subtract off means of MFCCs to normalize first

GMM Supervector Kernel

- MAP Adaptation of Means of UBM for each utterance is data point
- Compares parameters from unknown speaker to parameters from target and background speakers

$$K(utt_a, utt_b) = \sum_{i=1}^{N} w_i \mu_i^a \Sigma_i^{-1} \mu_i^b \tag{39}$$

- Current baseline SVM system for Speaker Verification (GMM-SVM based system)

Nuisance Attribute Projection

- Tries to project out subspaces that cause varability in the data.
- Nonlinear expansion of the GMM Supervector Kernel
- Compared to Factor Analysis, does not estimate variability

Table of Contents

Review of Speaker Verification

Introduction to Support Vector Machines

Derivation of SVM Equations

Soft Margin

Nonlinear Classification

SVMs in Speaker Verification

- Examples of Kernels Used in Speaker Verification
- How to Perform

How to Perform

- New SVM Decision boundary for each target speaker
- For MFCC-based kernels, target data are target MFCCs
- For GMM Supervector-based kernels, target data are adapted mean vectors from a UBM
- Background data are either background MFCCs or background adapted means

Sources for Pictures

- Slide 6: ZackWeinberg on Wikipedia adapting a picture by Cyc
- Slides 10, 12, 13, 14, 15: Cyc on Wikipedia
- Slides 18, 19, 20, 21: EMILeA-stat by Institut fr Statistik und Wirtschaftsmathematik (`http://emilea-stat.stochastik.rwth-aachen.de/cgi-bin/WebObjects/EMILeAstat.woa/wo/0.0.27.1.1.3.0`)
- Slide 24: Alisneaky on Wikipedia
- Slide 25: Tristan Fletcher, Support Vector Machines Explained, UCL. London, England