PSS718 - Data Mining

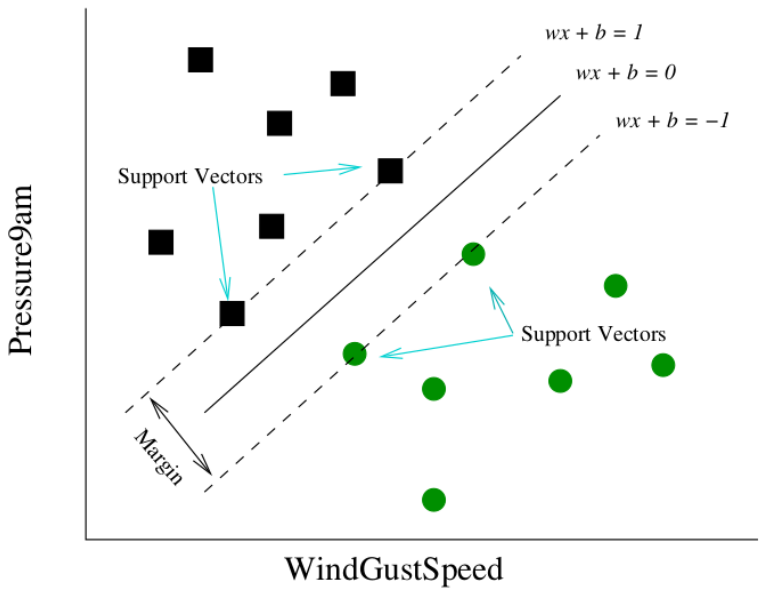
Lecture 11 - Support Vector Machines

Asst.Prof.Dr. Burkay Genç

Hacettepe University

December 11, 2016

What is it?



Definition (Support Vector)

An observation on the boundary of a class

Definition (Margin)

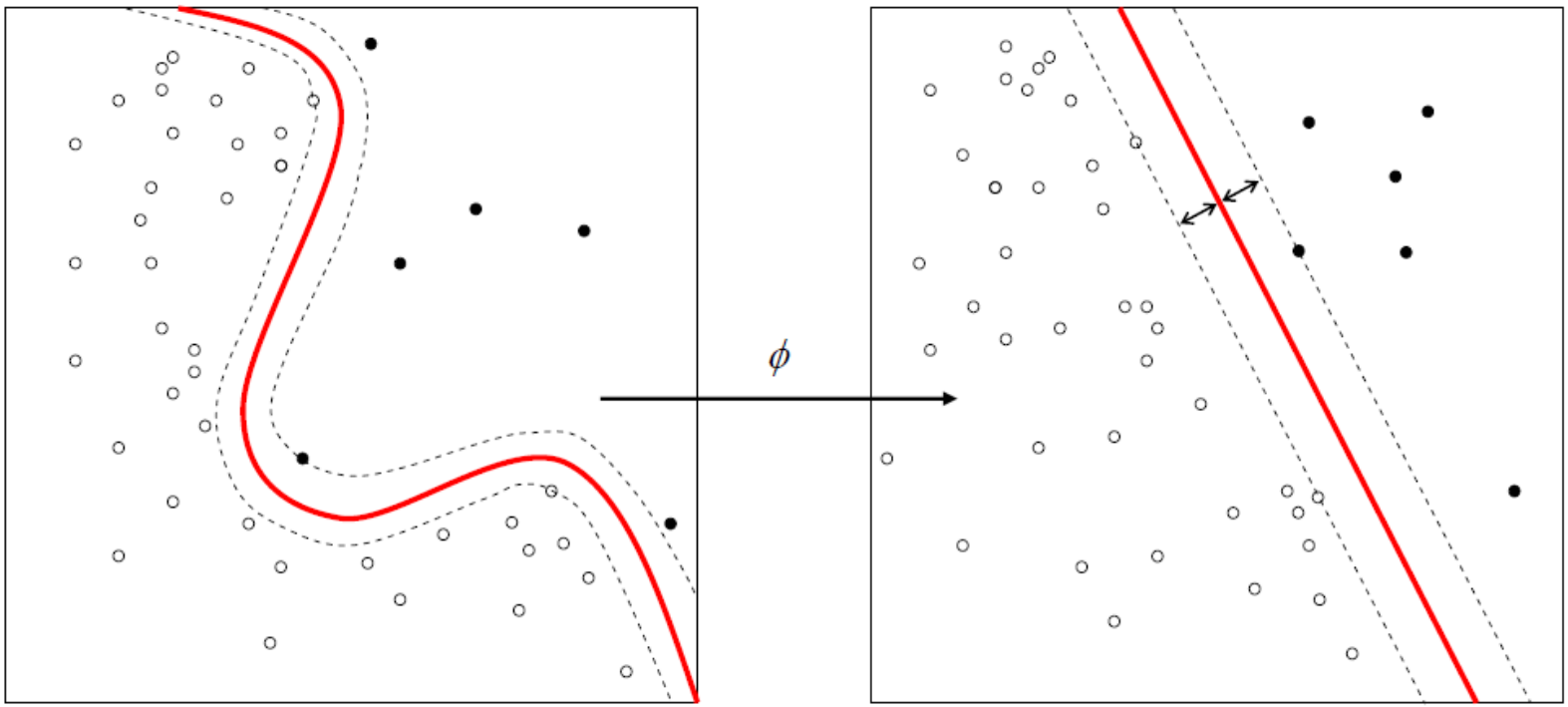
The distance between two class regions

Definition (Support Vector Machine)

A model describing support vectors of each class with maximum possible margin.

Are classes separable?

- Classes won't always be nicely separable with a straight line (or a corresponding hyperplane in higher dimensions)
- We then remap the data in different ways, creating new variables so that the classes become separable



Prediction

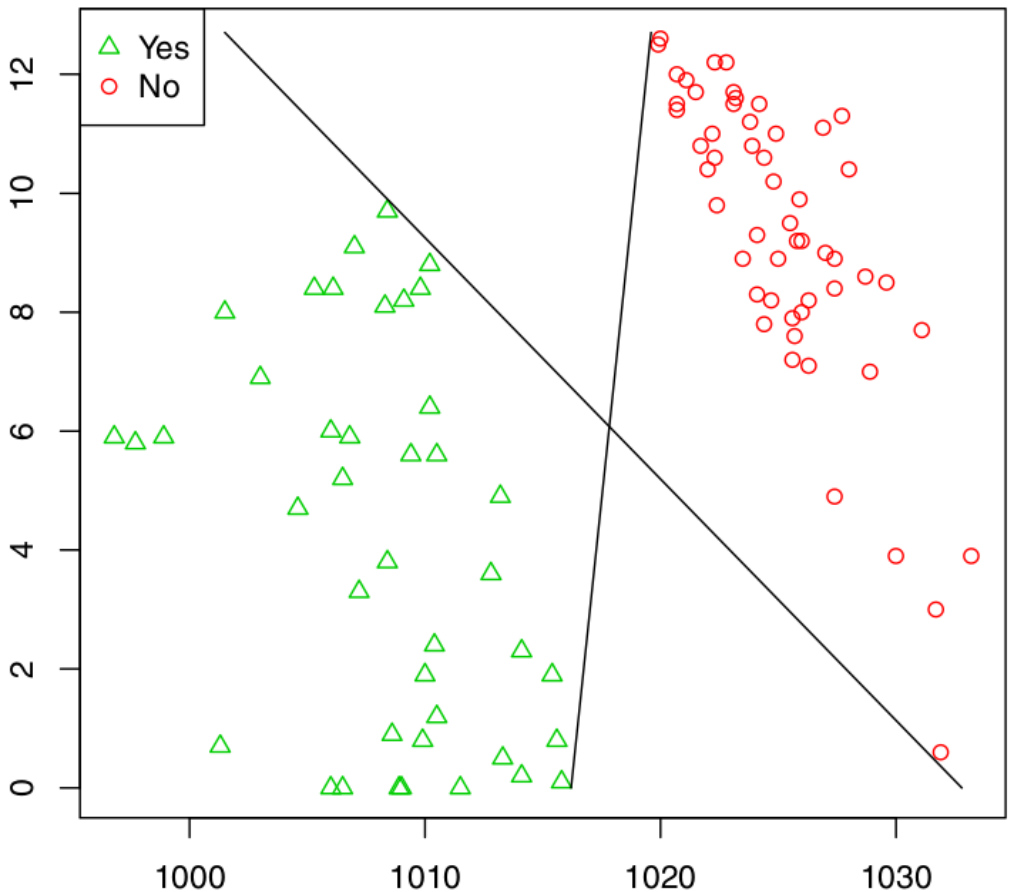
- Input X
- Remap X -> X'
- Check X' on map

Pros and Cons

- Good for
  - Nonlinear
  - Sparse
  - High-dimensional
  - Not sensitive to outliers
- Bad for
  - Sensitive to the type of transformations
  - Hard to use
  - Transformations may be expensive
  - Time consuming to find the best model

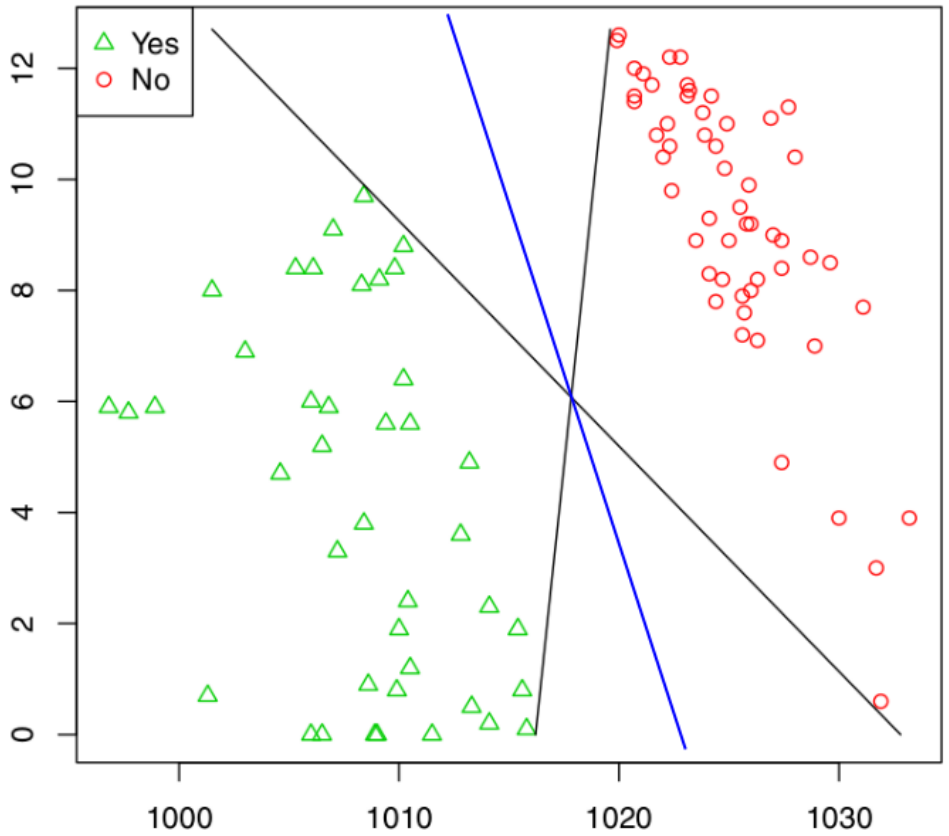
Main block

- This is a linear model
- We try to identify a line, plane...
- But there may be an infinite number of such lines



Confidence

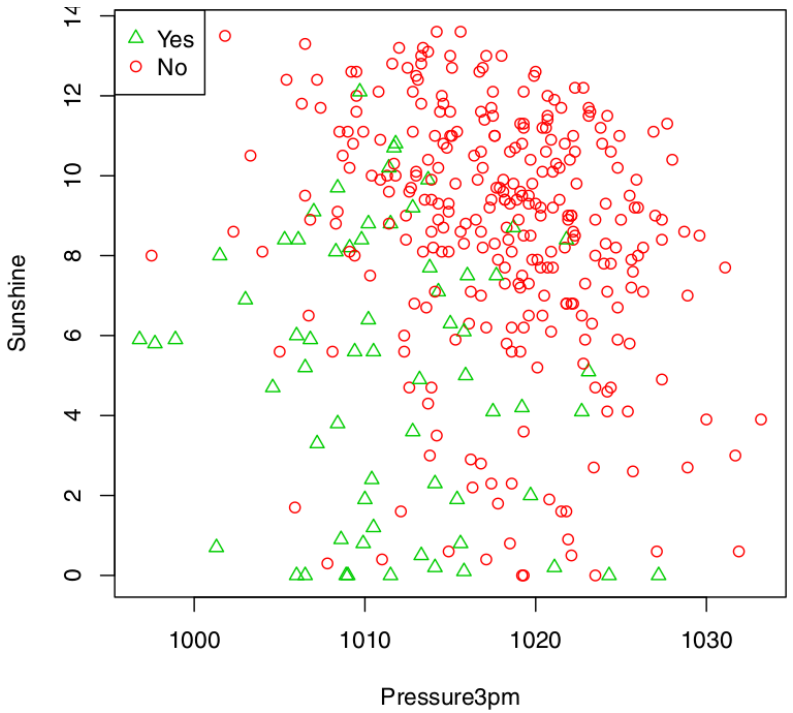
- The lines in the previous diagram are not very confident on the support vectors
- The line below is much more confident on them



Separability

- It is rarely the case that our observations are linearly separable
- Weather data example

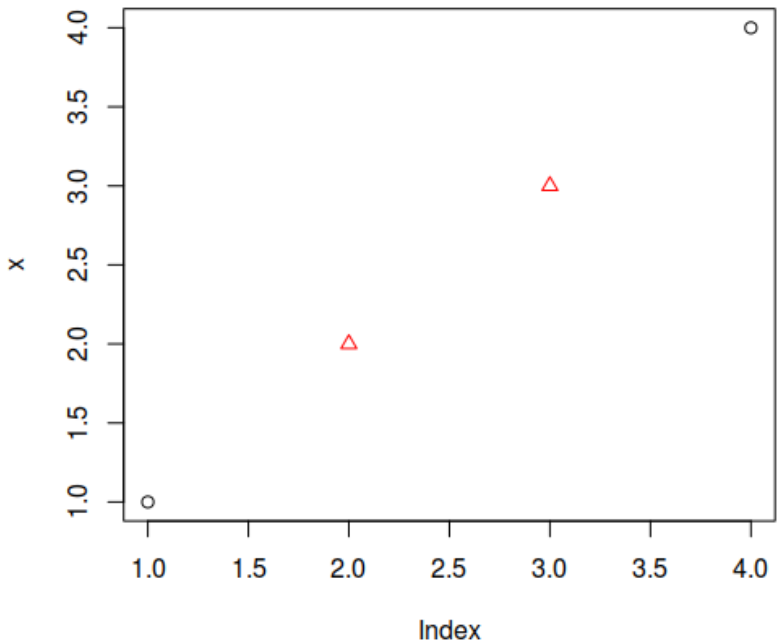
```
plot(Sunshine ~ Pressure3pm, data = weather,
col=1+as.numeric(RainTomorrow),
pch=as.numeric(RainTomorrow))
```

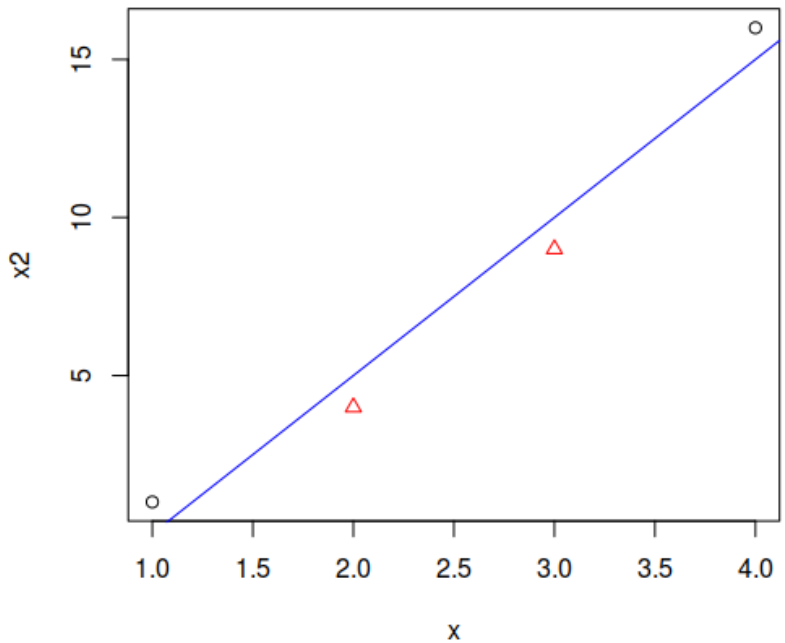


Transformations

- Some common types of transformations
  - Polynomial (homogeneous) : $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j)^d$
  - Polynomial (inhomogeneous) : $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d$
  - Gaussian radial basis function : $k(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2)$
  - Hyperbolic tangent : $k(\vec{x}_i, \vec{x}_j) = \tanh(\kappa \vec{x}_i \cdot \vec{x}_j + c)$

A VERY Simple Example






```
par(mfrow=c(1,2))
df <- data.frame(x = c(1,2,3,4), t = c(1,2,2,1))
plot(x~1, data=df, col=t, pch=t)
df <- cbind(df, x2 = df$x^2)
plot(x2~x, data=df, col=t, pch=t)
abline(lm(df$x2~df$x), col="blue")
```

Rattle support

- Rattle uses the kernlab package
  - Extensive collection of kernel methods
- kernlab has many tuning parameters
  - Only some are supported in Rattle

Output

```
Summary of the SVM model (built using ksvm):

Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.0394657153475283

Number of Support Vectors : 106

Objective Function Value : -59.2538
Training error : 0.100877
Probability model included.

Time taken: 0.07 secs
```

Radial Basis

- Requires two parameters C and sigma
- C is cost parameter and by default it is 1
- sigma is automatically computed by the algorithm