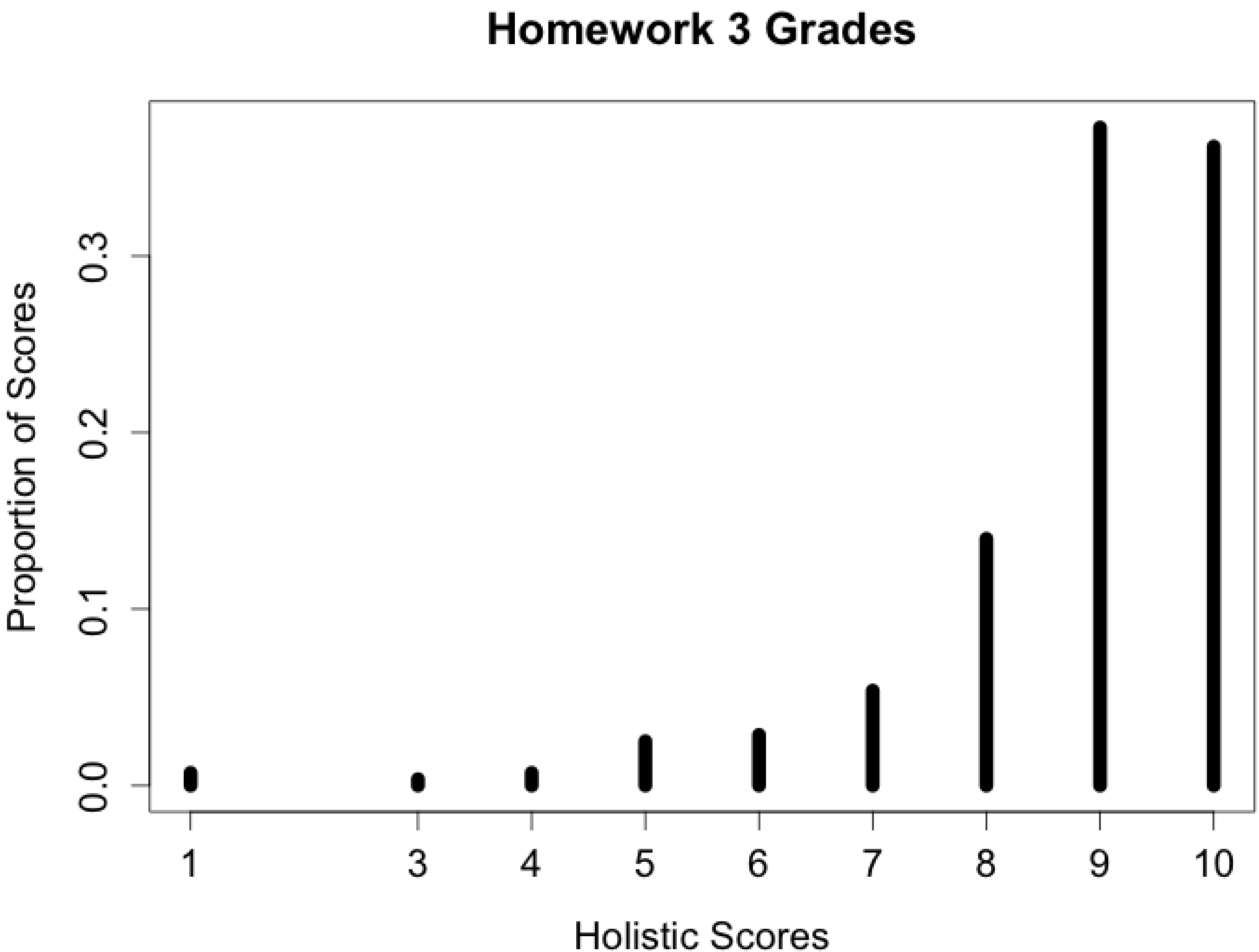
CS109 – Data Science

Verena Kaynig-Fittkau

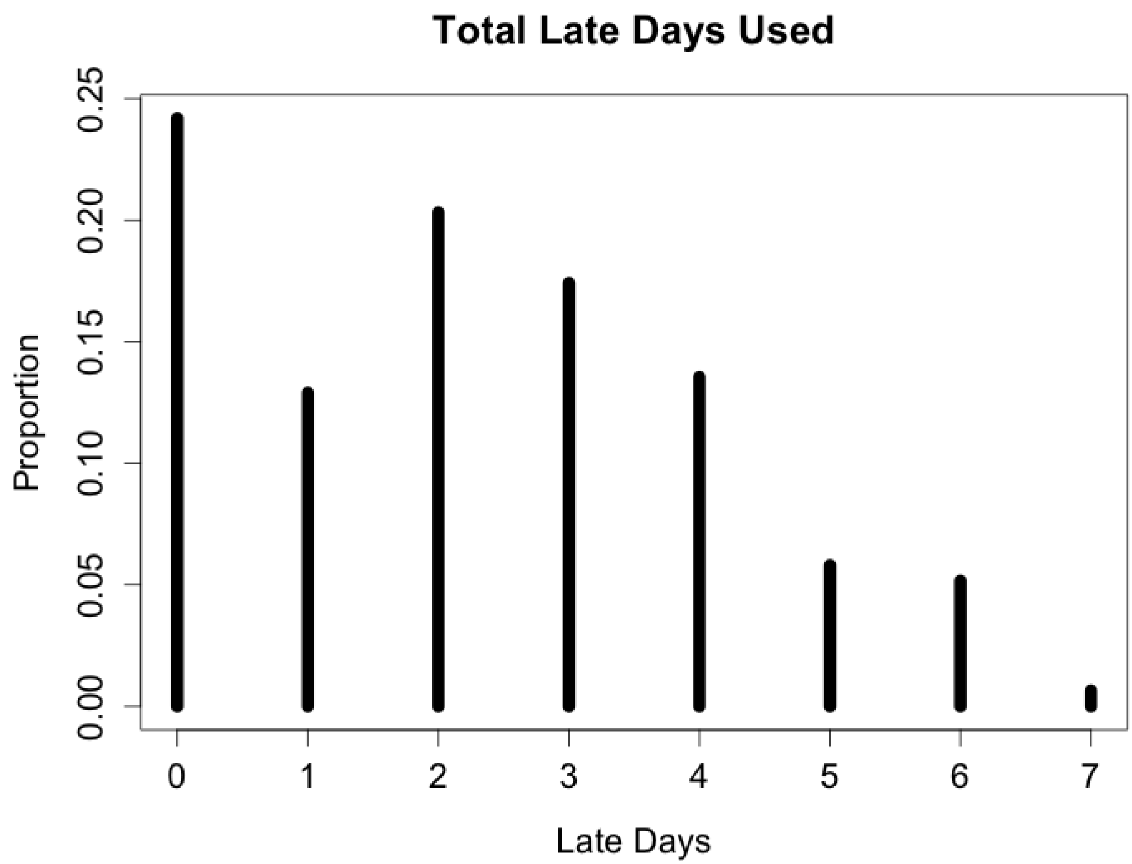
vkaynig@seas.harvard.edu
staff@cs109.org

Announcements



Announcements

- HW late days: 6 total!
- Afterwards we have to start to deduct points.
- Late days are reported in your HW comments email



Poll Competition

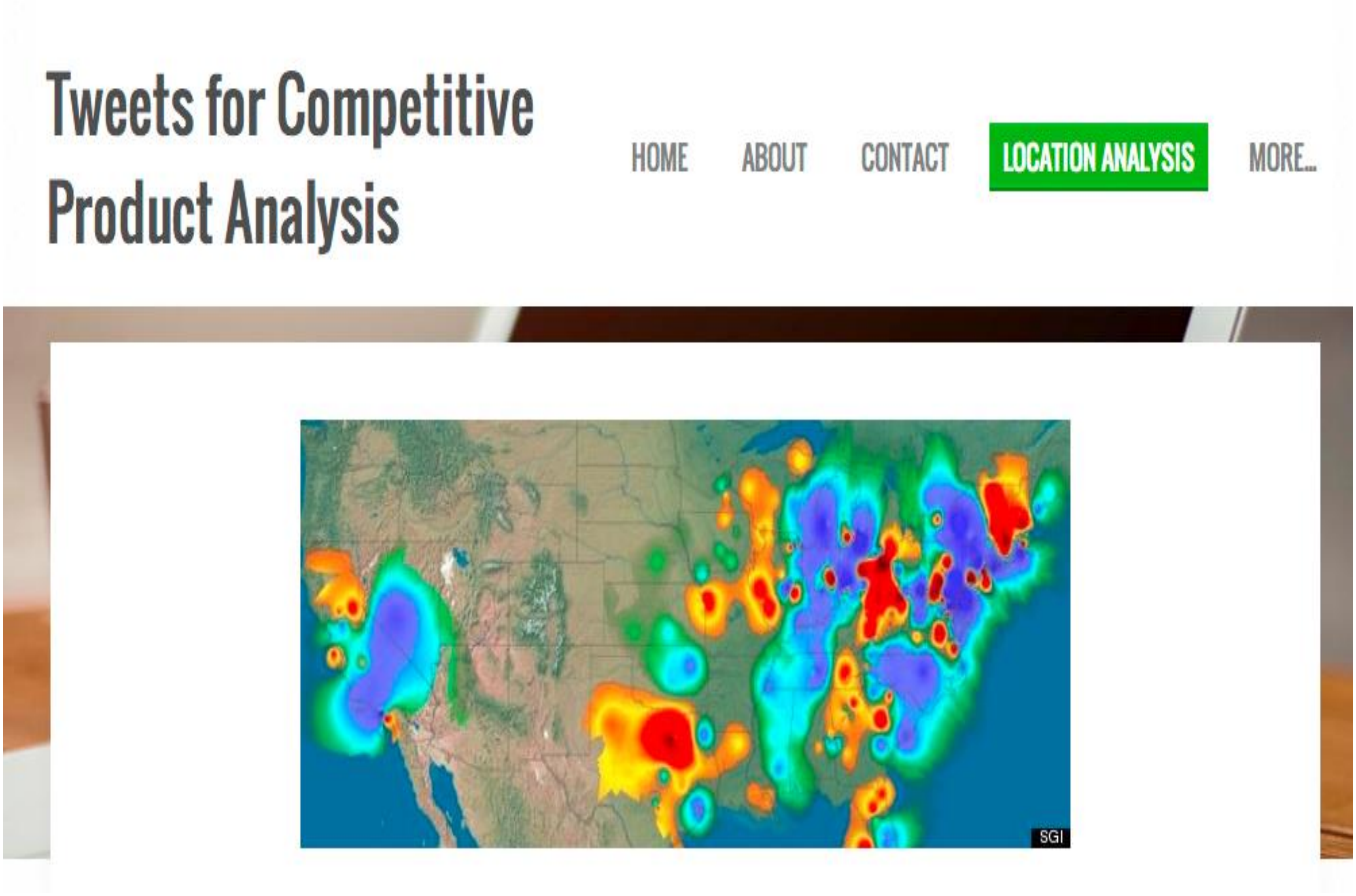
- Predict the midterms
- Just for fun!
- Form is on Piazza

Final Project

- Project description linked on the course homepage.
- Project proposals are due 11/17
- Week of 11/17: meet with your TF
- Projects due 12/10: Ipython notebook
- Website and video due: 12/12

IPython Process Book

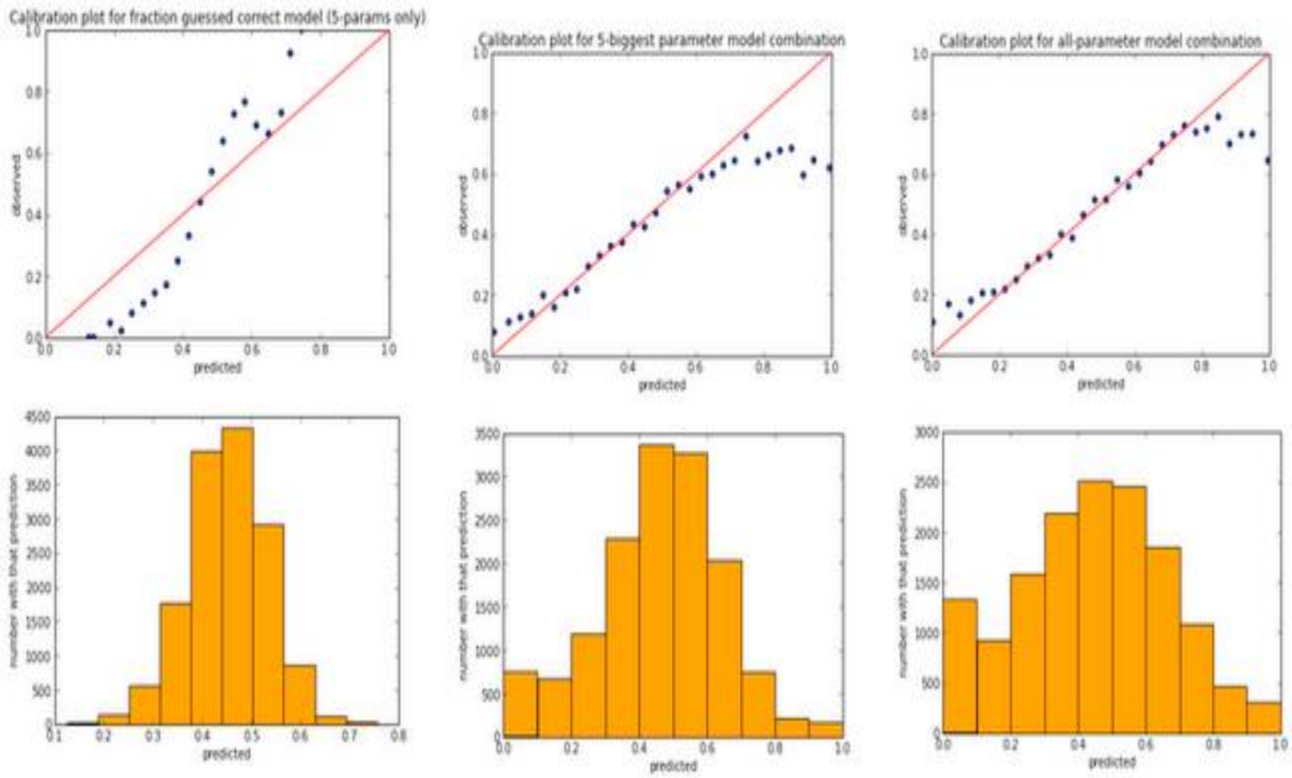
- Overview and Motivation
- Related Work
- Initial Questions
- Data
- Exploratory Analysis
- Final Analysis



http://youtu.be/w7kXX0SJcYk

Video

cs109.Kickstarter



https://www.youtube.com/watch?v=owYIbU_R5yA

Video

The Evolution of the American Presidency

https://www.youtube.com/watch?v=Z_ucSP5Fmdk#t=0

Kathy Lin, Renzo Lucioni, Matthew Moellman and Sherrie Wang

Video

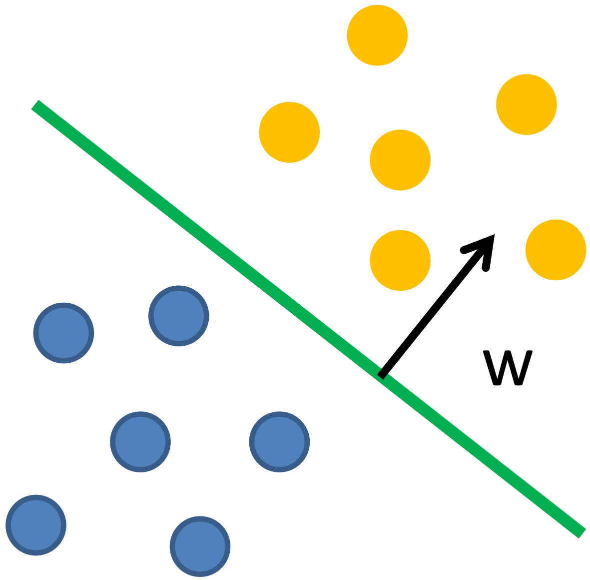
Take Home Messages

- Say why you are doing something
- It’s better to go safe than sorry
- If you think something should have worked and it didn’t, tell us what you think is the reason.

- Tell a progress story in your notebook
- Result report in the video

Separating Hyperplane

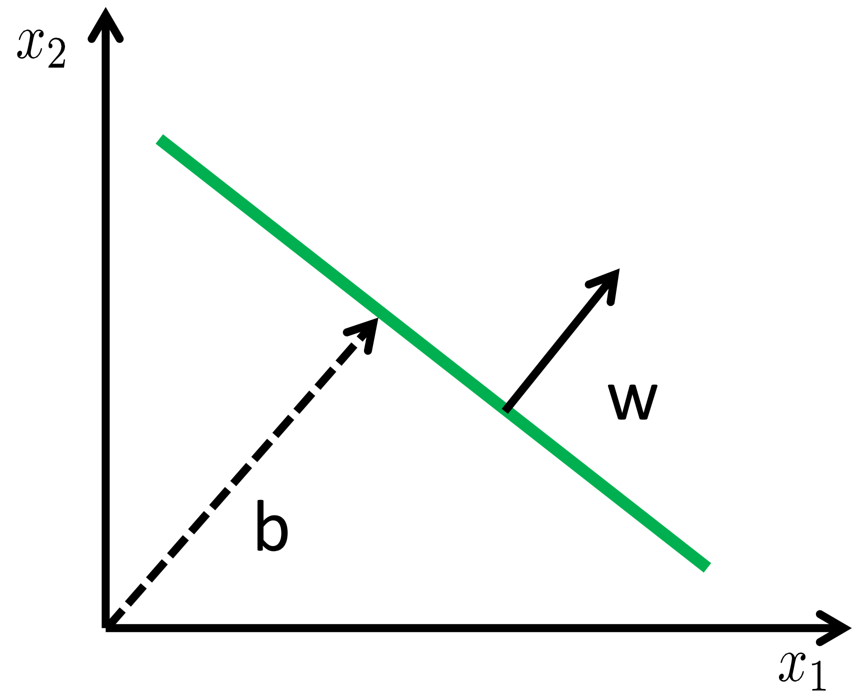
- x: data point
- y: label $\in \{-1, +1\}$
- w: weight vector
- b: bias



$$w^T x \qquad = 0$$

Separating Hyperplane

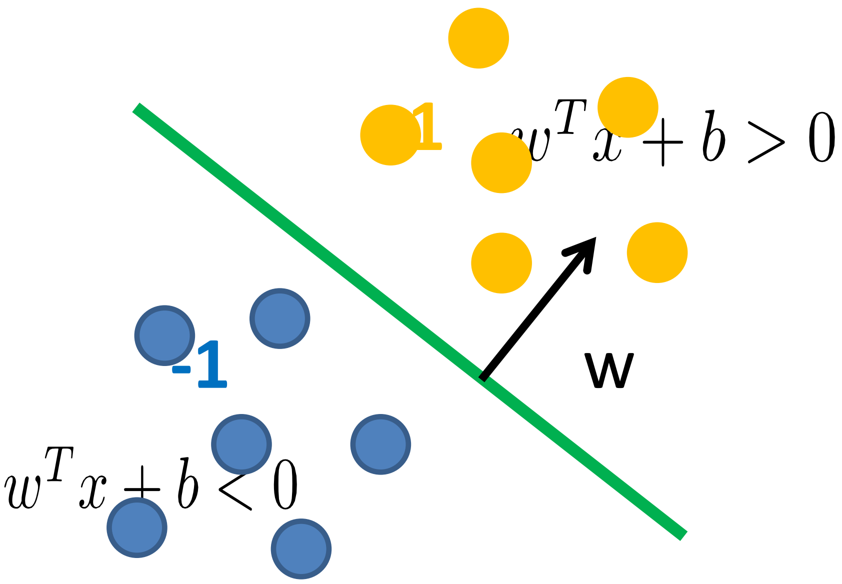
- x: data point
- y: label $\in \{-1, +1\}$
- w: weight vector
- b: bias



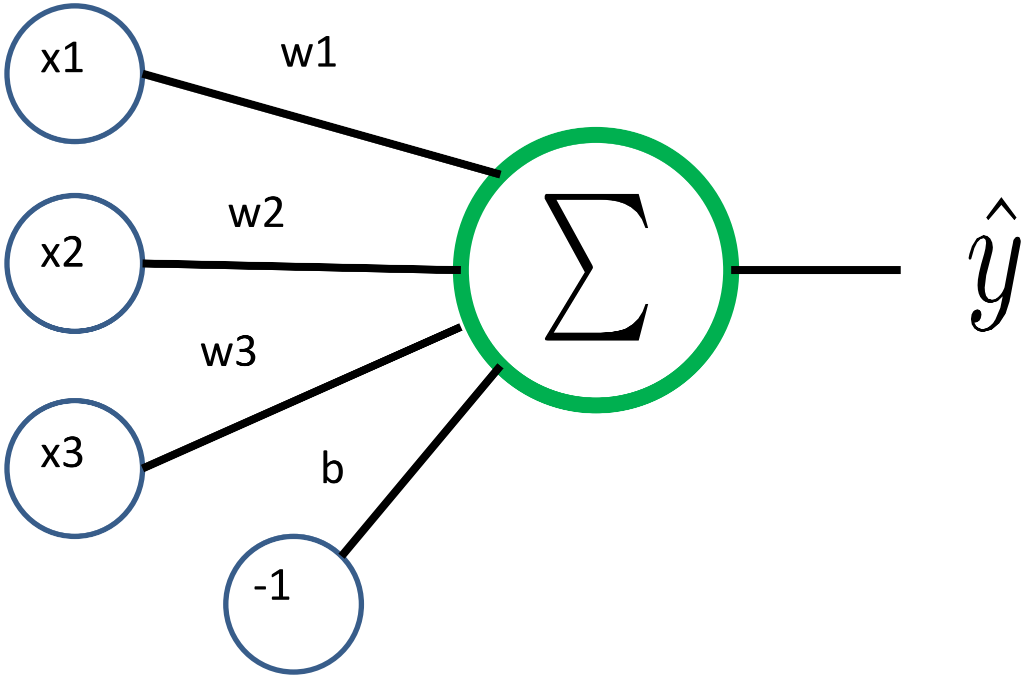
$$w^T x + b = 0$$

Separating Hyperplane

- x: data point
- y: label $\in \{-1, +1\}$
- w: weight vector
- b: bias

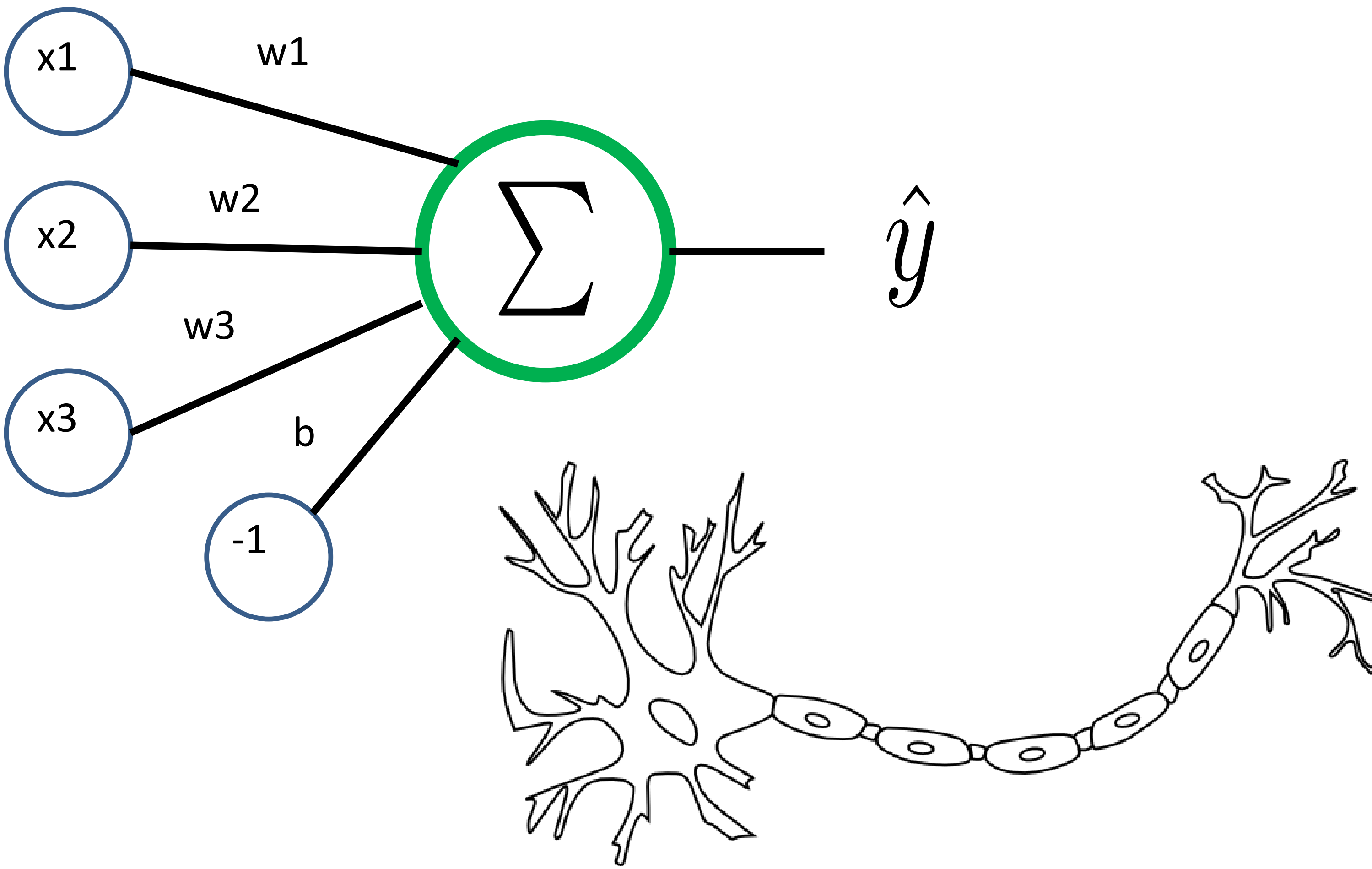


Perceptron



$$w^T x + b = 0$$

Perceptron Fun Facts



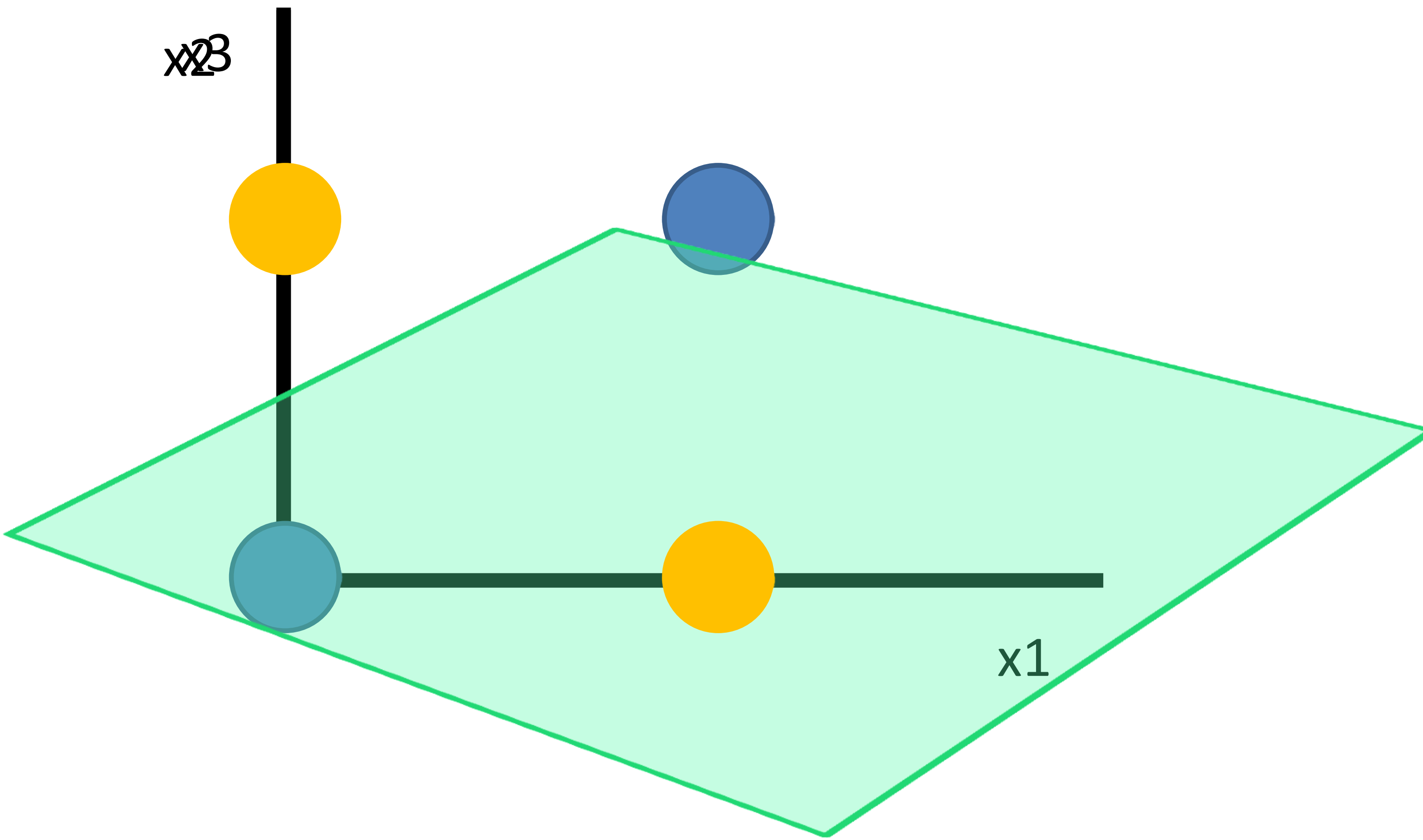
Perceptron Fun Facts

- invented 1957
- by Frank Rosenblatt

- the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence. (NYT 1958)

(http://en.wikipedia.org/wiki/Perceptron

The XOR Problem



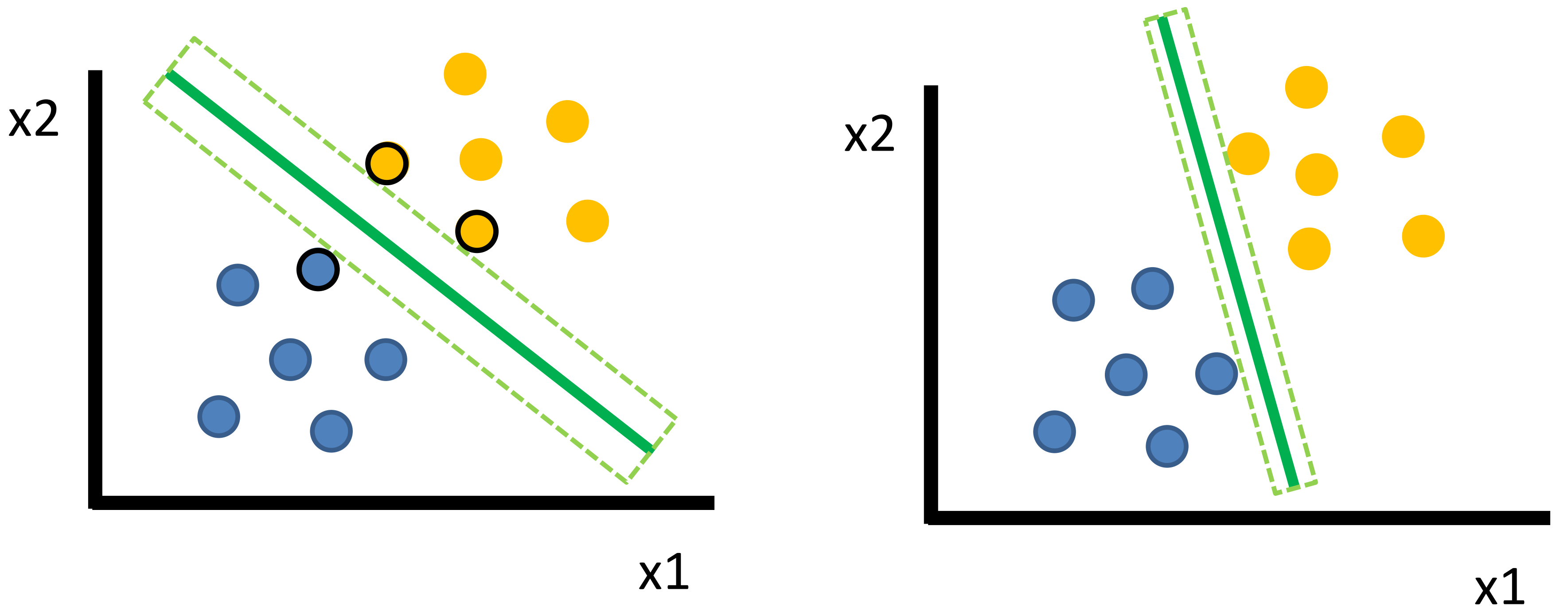
Support Vector Machine

- Widely used for all sorts of classification problems

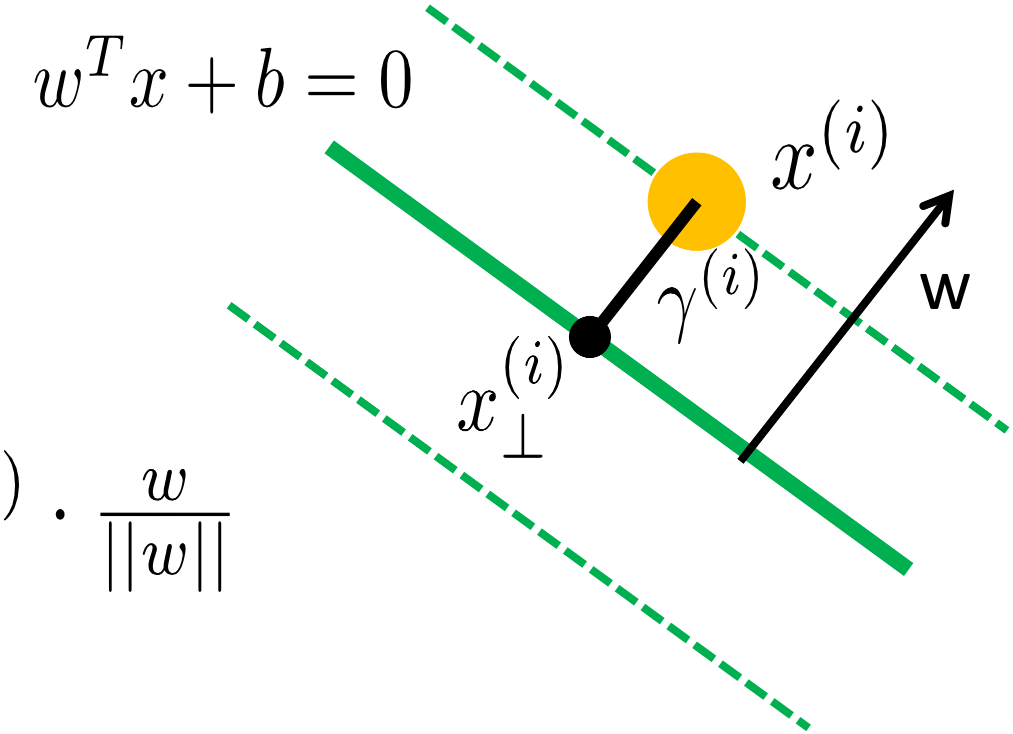
www.clopinet.com/isabelle/Projects/SVM/applist.html

- Some people say it is the best of the shelf classifier out there

Maximum Margin Classification



Maximum Margin Classification



margin:

$$x_\perp^{(i)} = x^{(i)} - \gamma^{(i)} \cdot \frac{w}{||w||}$$

$$w^T x_\perp^{(i)} + b = 0$$

$$\gamma^{(i)} = \quad \left(\frac{w^T x^{(i)} + b}{||w||} \right)$$

Maximum Margin Classification

$$\gamma^{(i)} = y^{(i)}\left(\frac{w^T x^{(i)} + b}{||w||}\right)$$ geometrical margin

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x + b)$$ functional margin

$$\max_{\gamma,w,b} \quad \gamma$$ minimal geometrical margin

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \quad i = 1, \ldots, m$$

$$||w|| = 1.$$ non-convex

Maximum Margin Classification

$$\begin{aligned} \max_{\gamma,w,b} \quad & \gamma \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \quad i = 1, \ldots, m \\ & ||w|| = 1. \end{aligned}$$

non-convex

$$\begin{aligned} \max_{\gamma,w,b} \quad & \frac{\hat{\gamma}}{||w||} \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, m \end{aligned}$$

Maximum Margin Classification

$$\max_{\gamma,w,b} \quad \frac{\hat{\gamma}}{||w||}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, m$$

functional margin is not normalized – can be arbitrarily scaled

$$\min_{\gamma,w,b} \quad \frac{1}{2}||w||^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, m$$

SVM Applet

http://www.ml.inf.ethz.ch/education/lectures_and_seminars/annex_estat/Classifier/JSVMLinearKernelApplet.html

What about outliers?

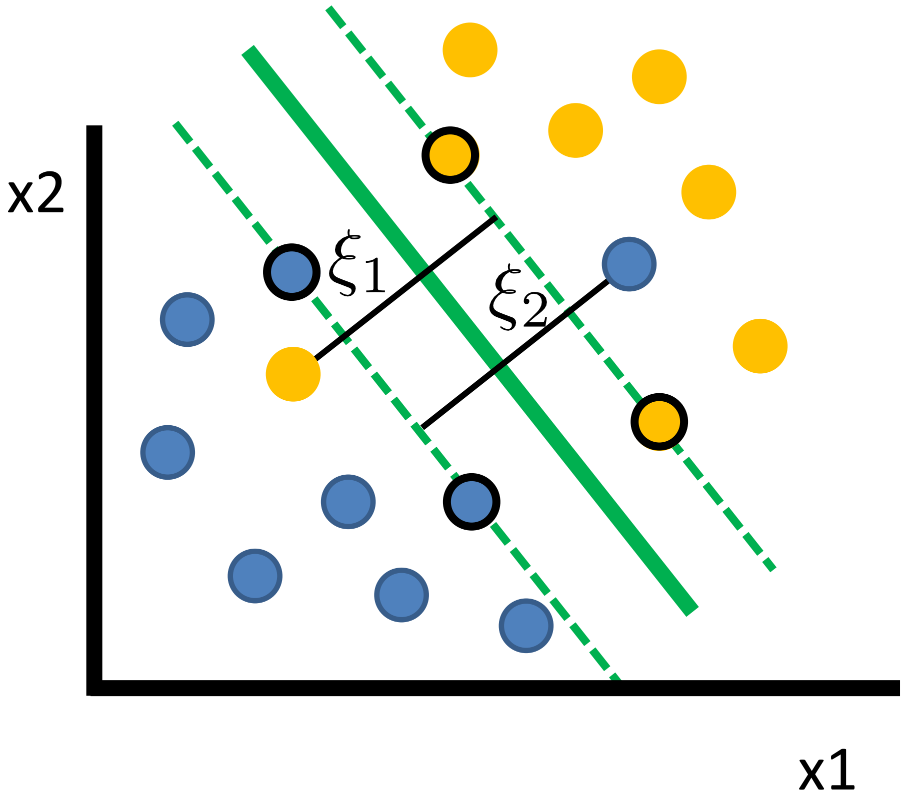
ξ_i: slack variables

$\min_{w,b,\xi} \frac{1}{2}||w||^2$

subject to:

$y^{(i)}(w^T x^{(i)} + b) \geq 1$

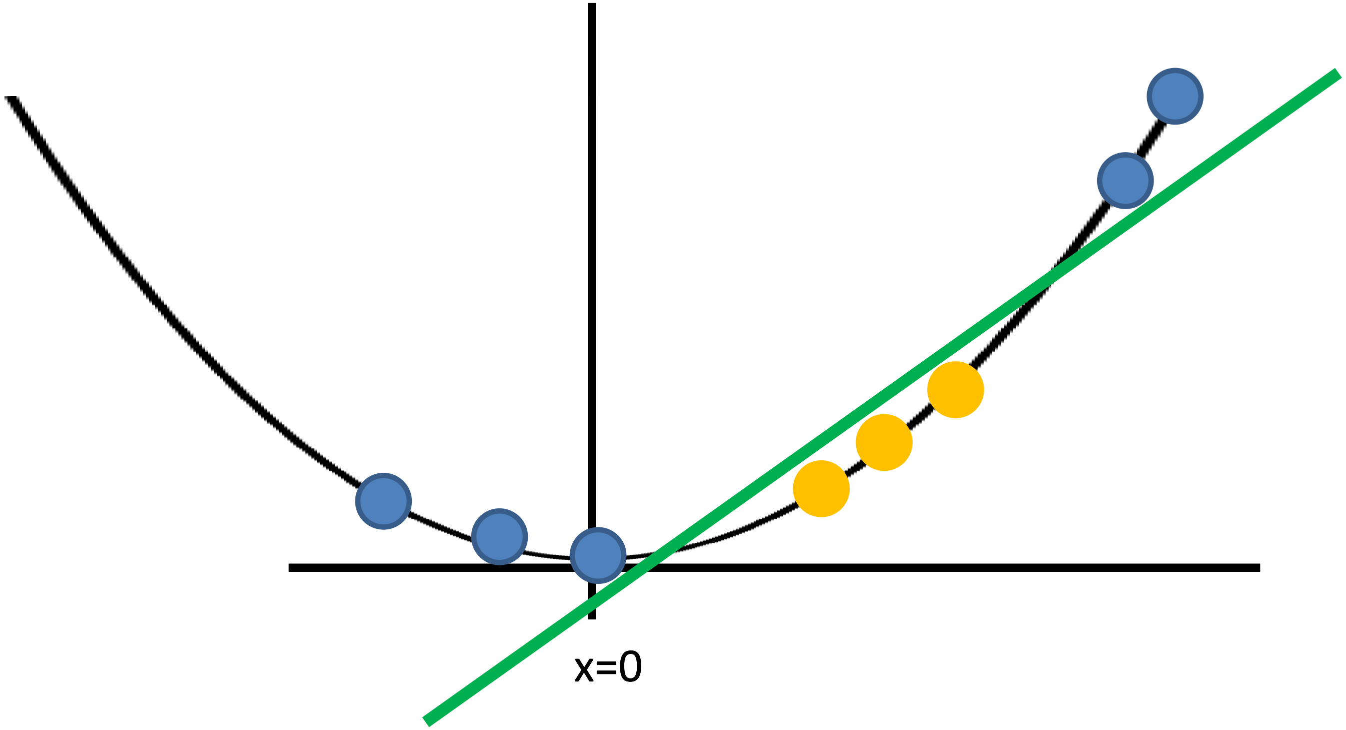
$(i = 1, \ldots, n)$



SVM Applet

http://www.ml.inf.ethz.ch/education/lectures_and_seminars/annex_estat/Classifier/JSVMLinearKernelApplet.html

XOR problem revised



Did we add information to make the problem seperable?

SVM with a polynomial Kernel visualization

Created by:
Udi Aharoni

Quadratic Kernel

$$x = (x_1, x_2)$$

$$\Phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

$$\Phi(x) \cdot \Phi(z) = 1 + 2\sum_{i=1}^{d} x_i z_i$$

$$+\sum_{i=1}^{d} x_i^2 z_i^2 + 2\sum_{i=1}^{d}\sum_{j=i+1}^{d} x_i x_j z_i z_j$$

$$\boxed{= (1 + x \cdot z)^2}$$

Kernel Functions

$$K(x,z) = \Phi(x) \cdot \Phi(z)$$

- Polynomial:

$$K(x,z) = (1 + x \cdot z)^s$$

- Radial basis function (RBF):

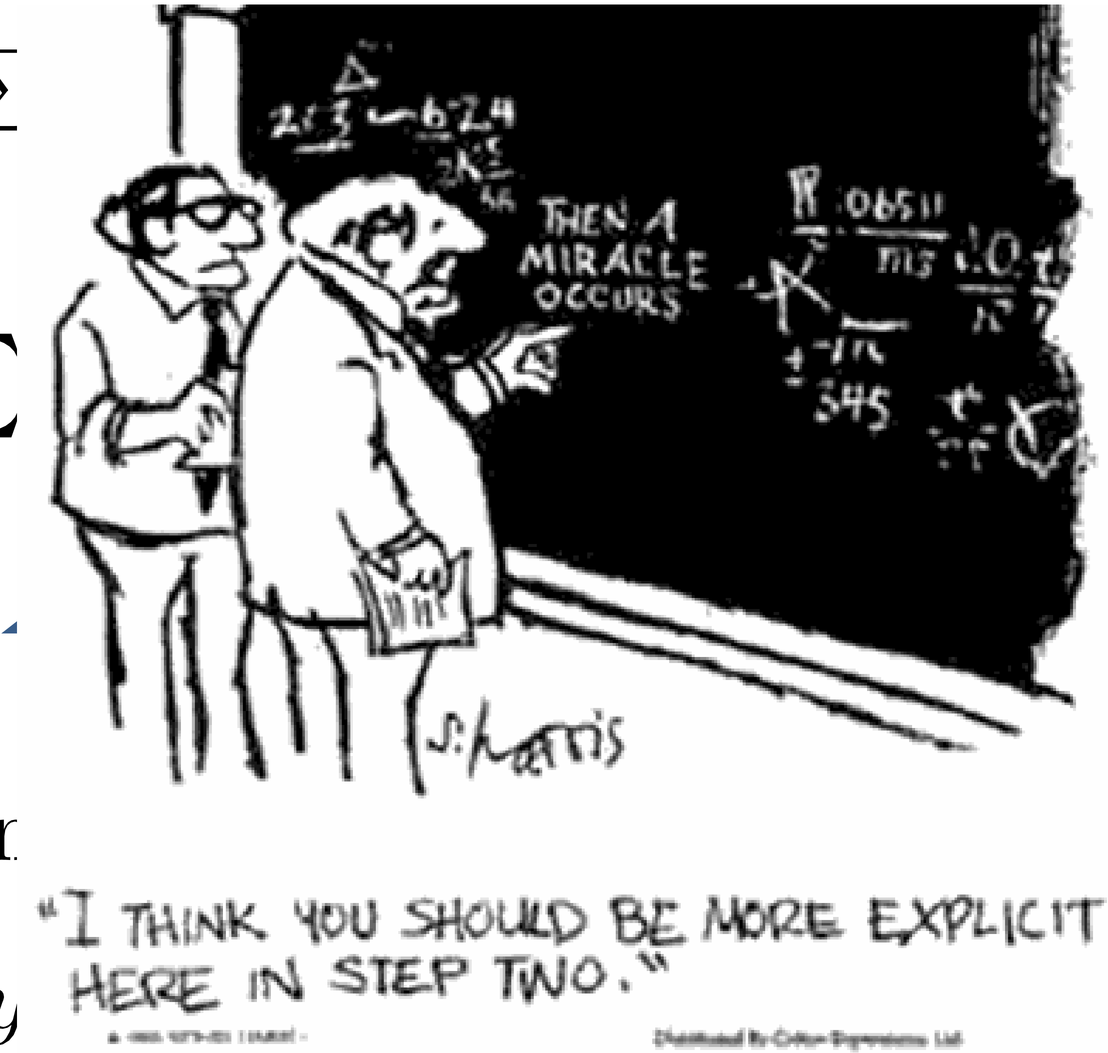
$$K(x,z) = \exp(-\gamma(x - z)^2)$$

So what is the excitement?

$\max_\alpha \sum$ $(i)^T x^{(j)}$

s.t. α_i

$\sum$



$\arg\mathrm{r}$

s.t. y

So what is the excitement?

$$\max_\alpha \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i,j=1}^m y^{(i)}y^{(j)}\alpha_i\alpha_j \boxed{{x^{(i)}}^T x^{(j)}}$$

$$\text{s.t. } \alpha_i \geq 0, \; i = 1, \ldots, m$$

$$\sum_{i=1}^m \alpha_i y^{(i)} = 0$$

$$\boxed{K(x^{(i)}, x^{(j)})}$$

$$\arg\min_{w,b} \frac{1}{2}||w||^2$$

$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1$$

The Miracle Explained

- Andrew Ng does this really well

- http://cs229.stanford.edu/notes/cs229-notes3.pdf
- Course is also on Youtube, ItunesU, etc.

Kernel Trick for SVMs

- Arbitrary many dimensions
- Little computational cost
- Maximal margin helps with curse of dimensionality

SVM Applet, Part 2

http://www.ml.inf.ethz.ch/education/lectures_and_seminars/annex_estat/Classifier/JSupportVectorApplet.html

Tips and Tricks

- SVMs are not scale invariant
- Check if your library normalizes by default
- Normalize your data
  - mean: 0 , std: 1
  - map to [0,1] or [-1,1]
- Normalize test set in same way!

Tips and Tricks

- RBF kernel is a good default
- For parameters try exponential sequences
- Read:

  Chih-Wei Hsu et al., "A Practical Guide to Support Vector Classification", Bioinformatics (2010)