CSCI 5622 Machine Learning

ML

Support Vector & Kernel Machines

DATE	READ	DUE
Today, Oct 5	Burgess & Cristianini	
Wed, Oct 7	7	Notes Papers 3&4
Mon, Oct 12	Bagging & Boosting	Exper. 1 plan (1 pg)

www.RodneyNielsen.com/teaching/CSCI5622-F09/

Instructor: Rodney Nielsen

Assistant Professor Adjunct, CU Dept. of Computer Science

Research Assistant Professor, DU, Dept. of Electrical & Computer Engr.

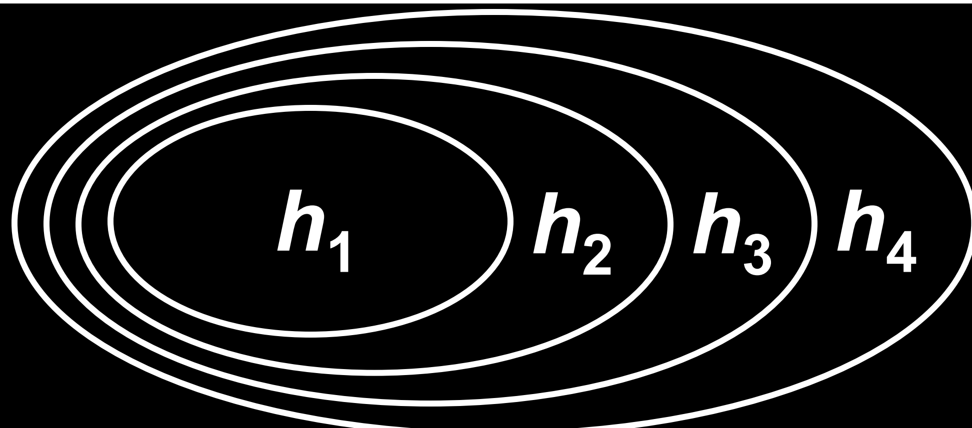
Research Scientist, Boulder Language Technologies

ML

Structural Risk Minimization

- **Find *h* that minimizes the actual risk**
- **Train a learner for each subset**
- **Choose min of empirical risk + VC confidence**

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(1 + \log 2N/h) - \log \eta/4}{N}}$$



ML

Support Vector Machines (SVMs)

- Generalization usually as good and often significantly better than other methods

ML

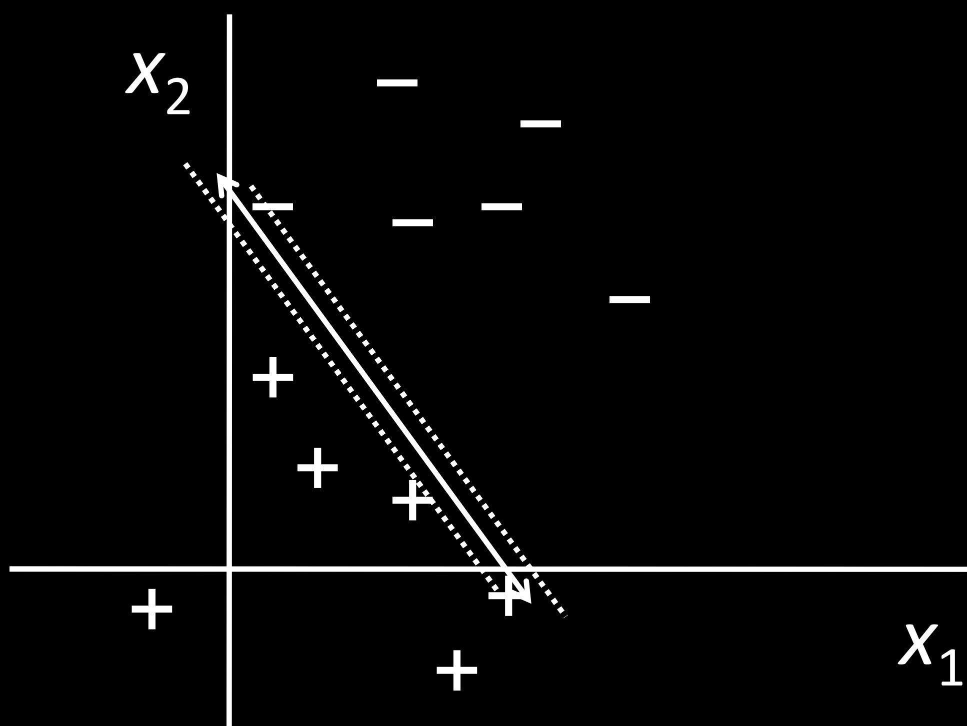
SVM Formulation

- **$\mathbf{wx}^{(i)} + b \geq +1$ for $y^{(i)} = +1$**
- **$\mathbf{wx}^{(i)} + b \leq -1$ for $y^{(i)} = -1$**
- **Or $y^{(i)}(\mathbf{wx}^{(i)} + b) - 1 \geq 0$ for all i**
- **Minimize $||\mathbf{w}||^2$, subject to above constraints (one for each training instance)**

ML

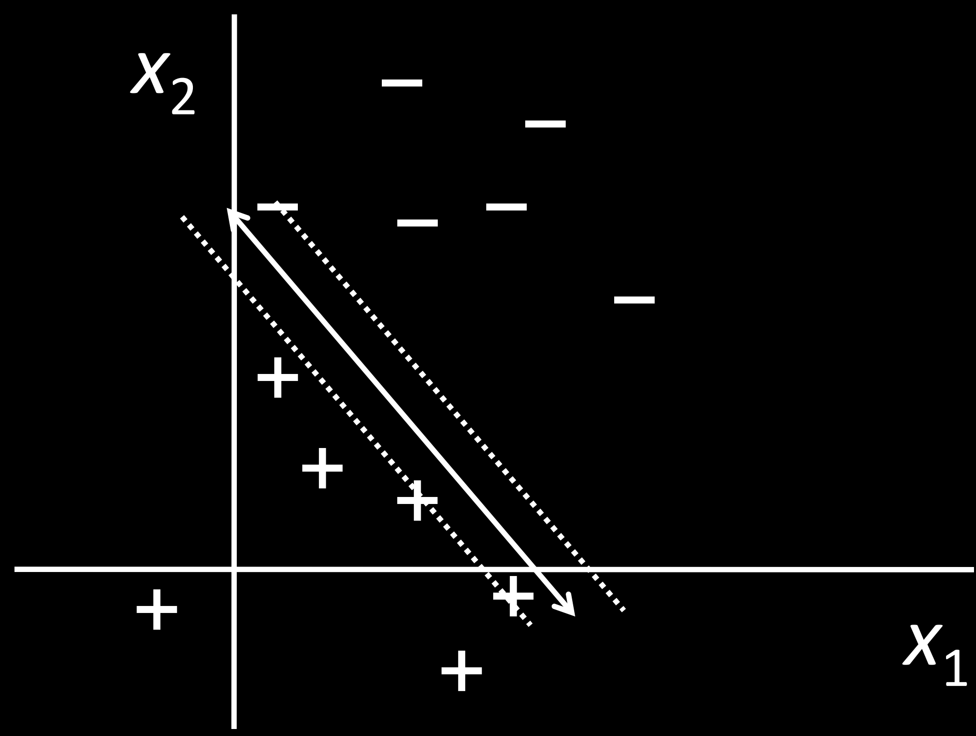
Linear Support Vector Machines

- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML Linear Support Vector Machines

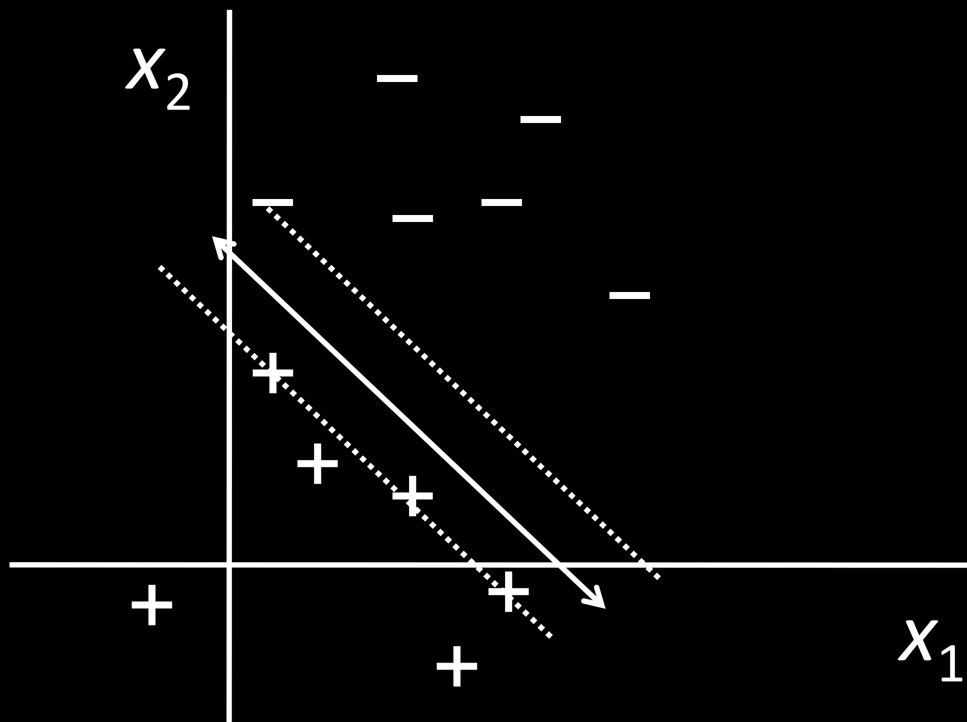
- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML

Linear Support Vector Machines

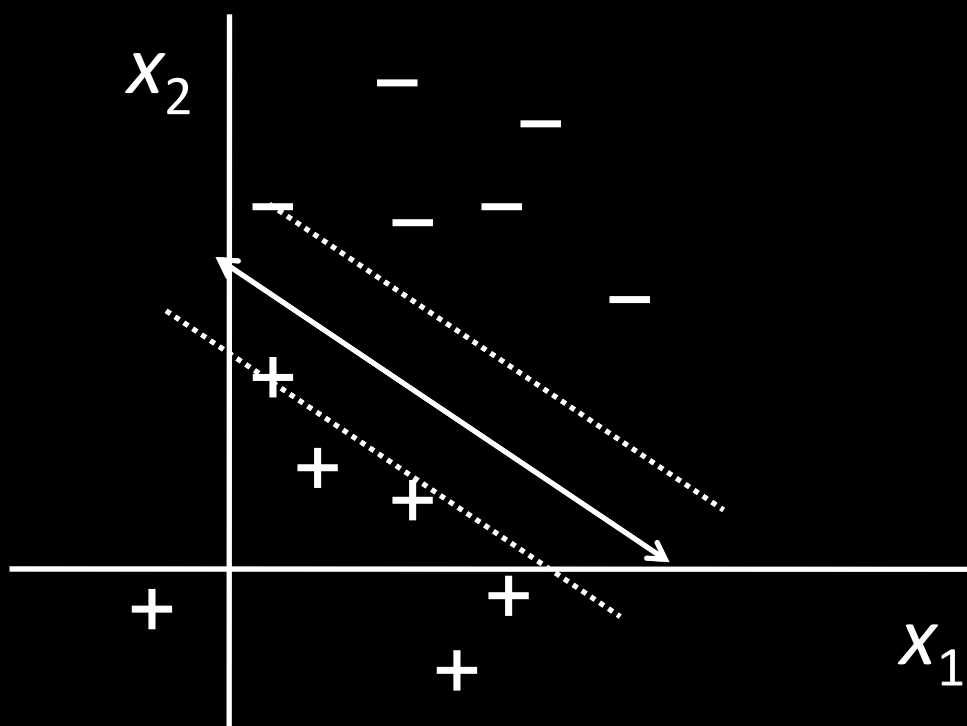
- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML

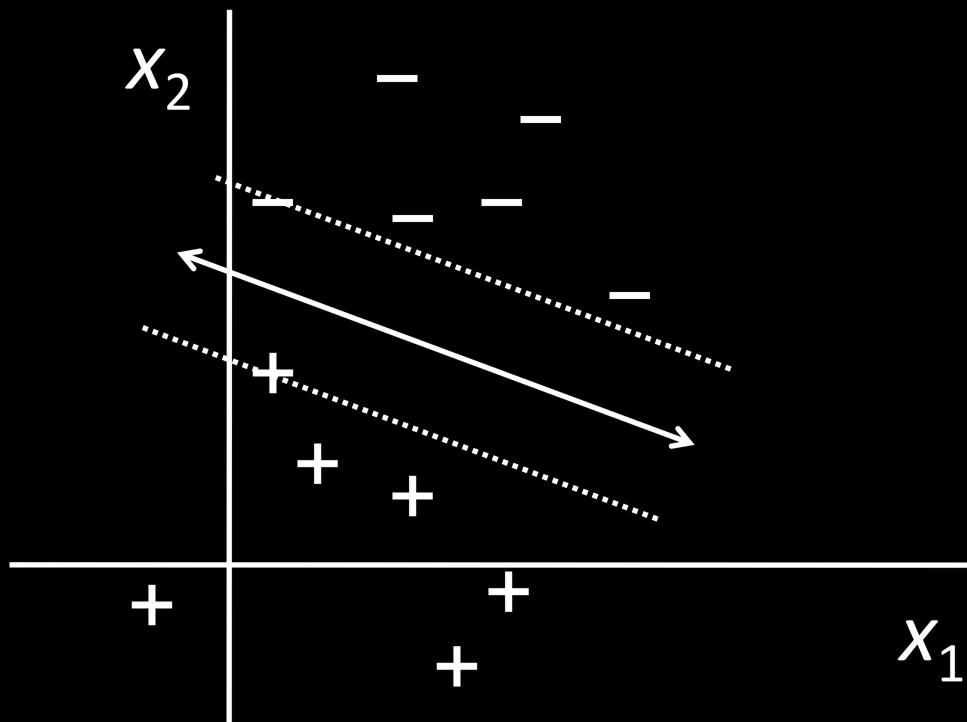
Linear Support Vector Machines

- Search for hyperplane that maximizes the margin ($d_+ + d_-$)



Linear Support Vector Machines

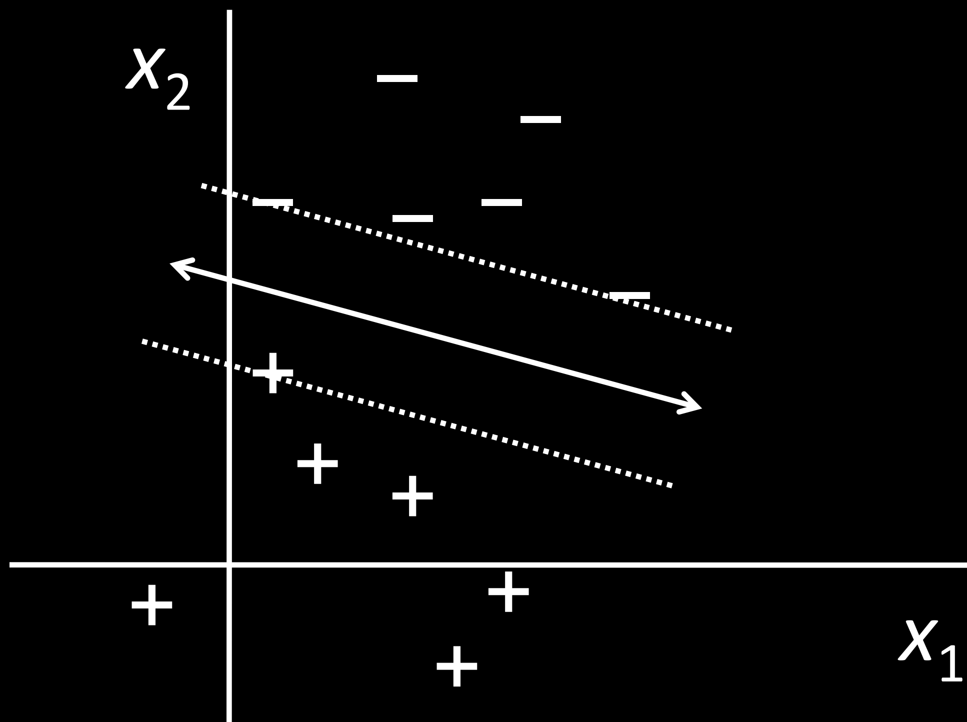
- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML

Linear Support Vector Machines

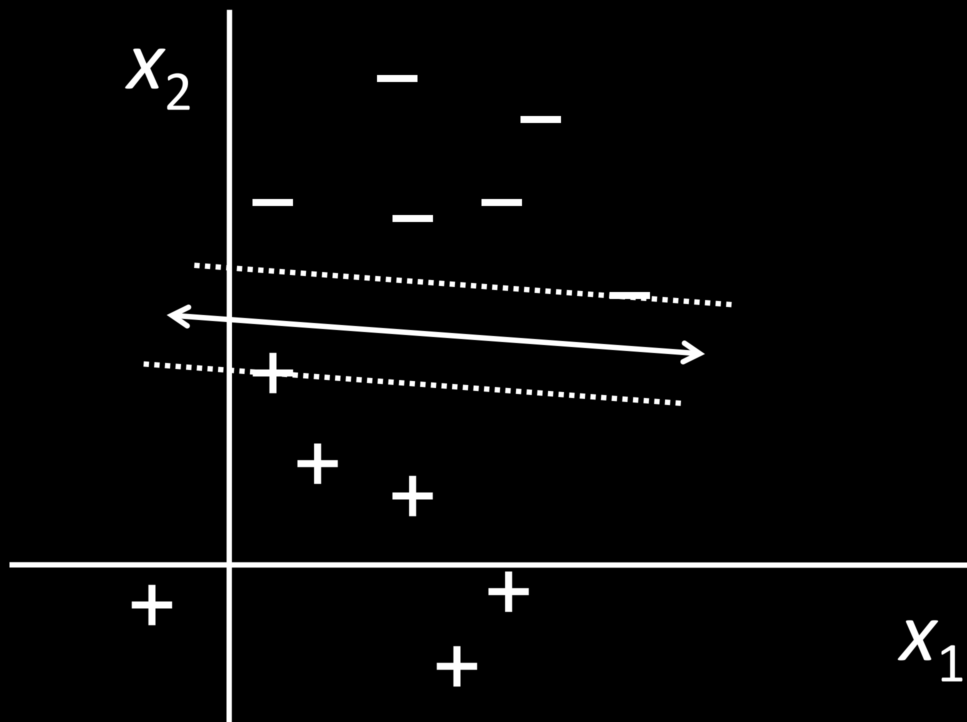
- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML

Linear Support Vector Machines

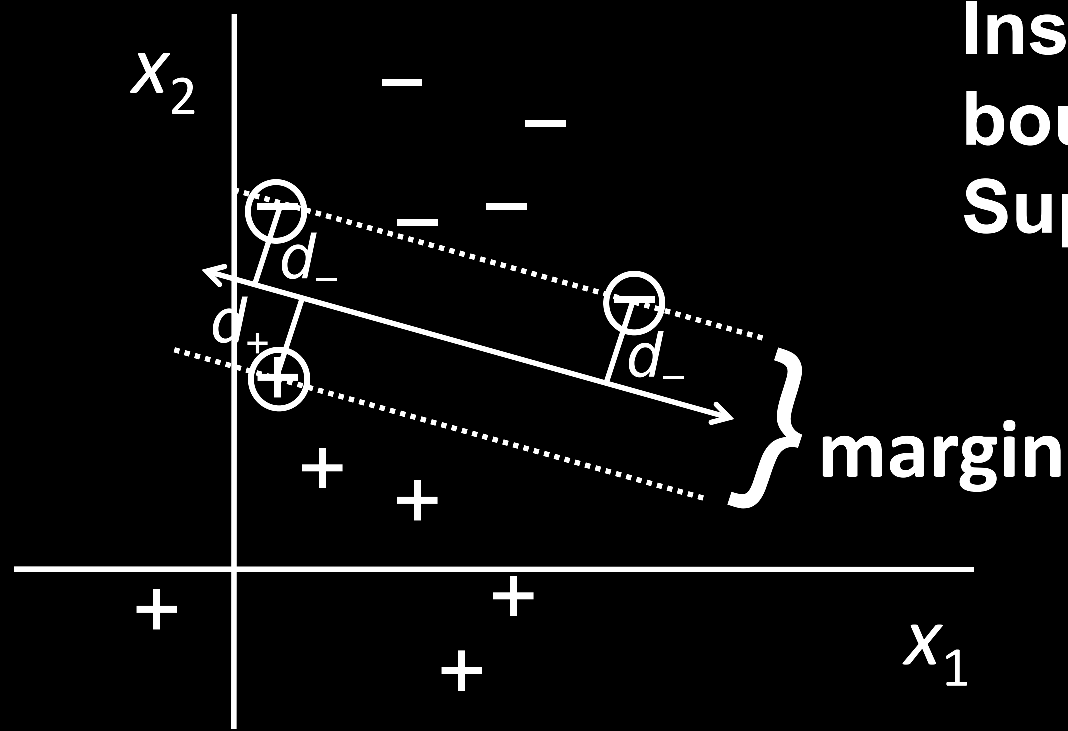
- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



ML

Linear Support Vector Machines

- **Search for hyperplane that maximizes the margin ($d_+ + d_-$)**



Instances on the boundaries are the Support Vectors

ML Support Vector Machines (SVMs)

- **SVMs are Linear Learning Machines represented in a dual fashion**

$$f(\mathbf{x}) = \operatorname{sgn}\left(\mathbf{w}^T\mathbf{x} + b\right) = \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} \mathbf{x}^{(i)T}\mathbf{x} + b\right)$$

$$\mathbf{w} = \sum_{i=1..N} \alpha_i y^{(i)} \mathbf{x}^{(i)}$$

ML

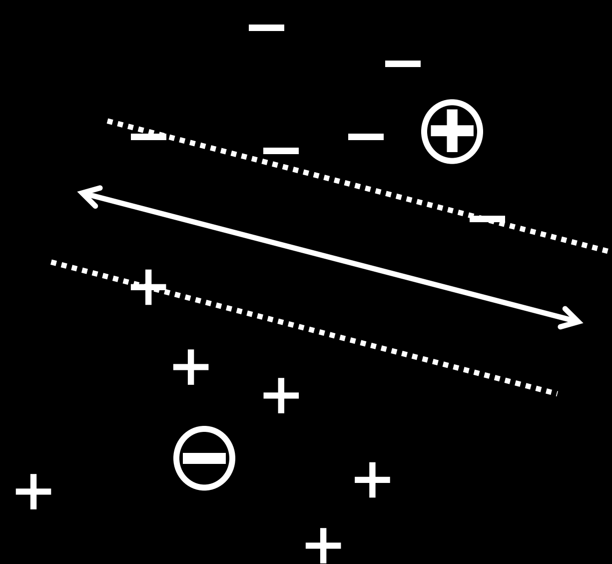
Classifier Decision Boundaries

- **Perceptron**
  - **Finds any separating hyperplane**
- **ANN**
  - **Finds a potentially complex, overfit, boundary**
- **SVM**
  - **Finds the maximum margin boundary for a linear hyperplane**

ML

Soft Margin Classification

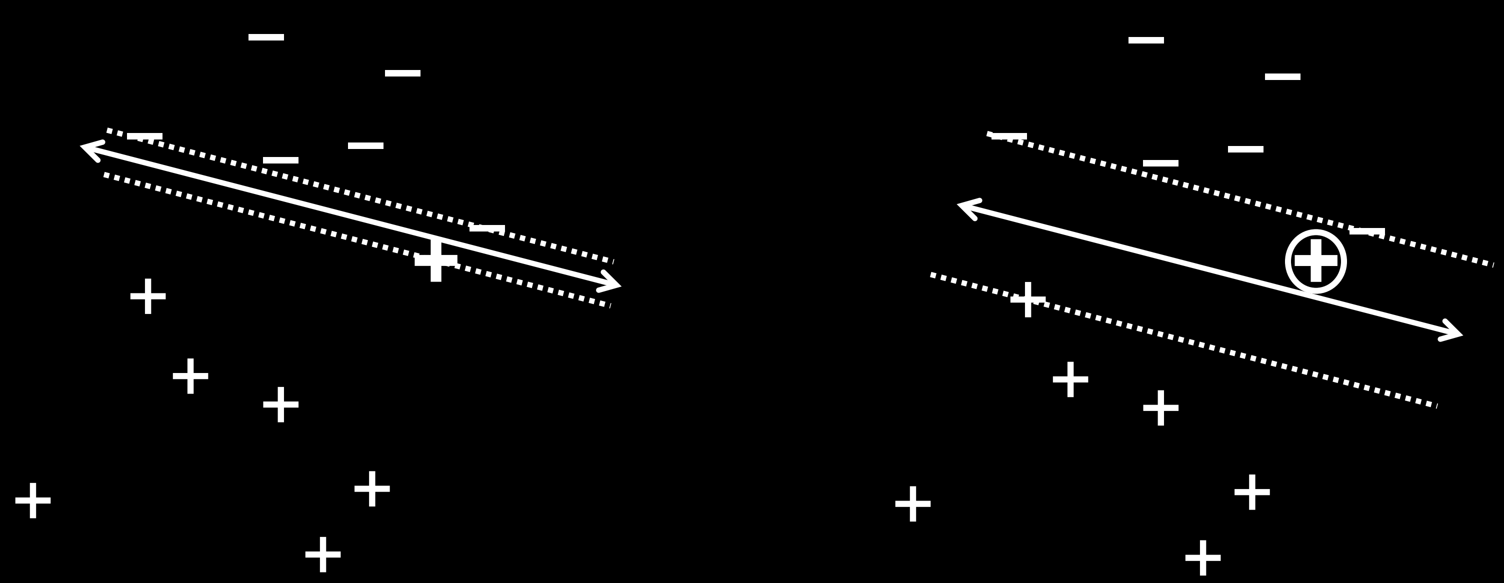
- **Introduce slack variables which effectively penalizes solutions for each error in the training data**



ML

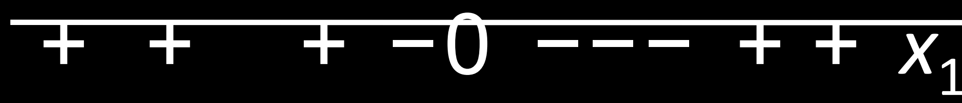
Soft Margin Classification

- **Introduce slack variables which effectively penalizes solutions for each error in the training data**

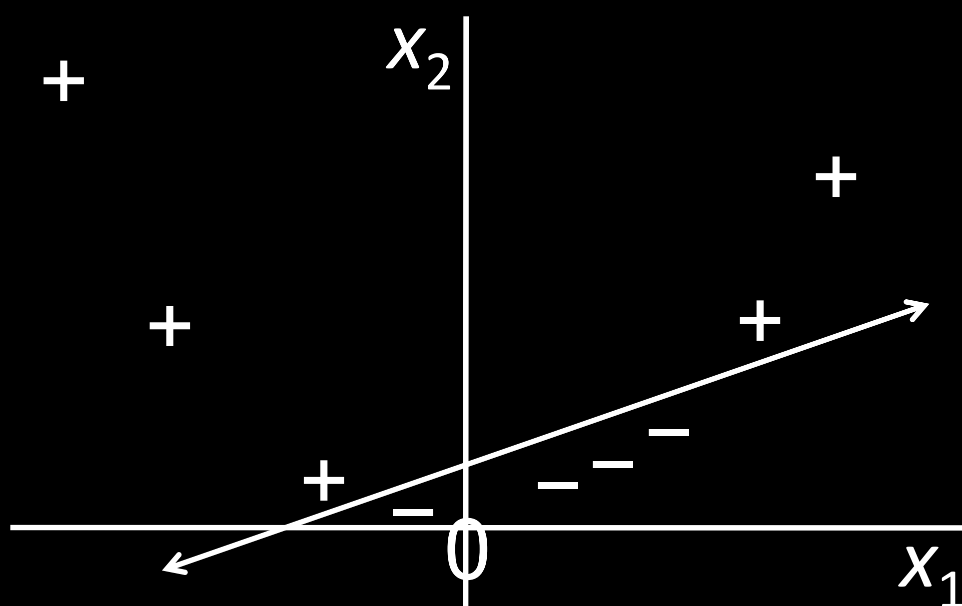


ML

Nonlinear SVMs



- **Project data to higher dimension**



Kernels

- **Kernels are an effective method of projection**

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} \mathbf{x}^{(i)^T} \mathbf{x} + b \right)$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} \phi\left(\mathbf{x}^{(i)}\right)^T \phi(\mathbf{x}) + b \right)$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} K\left(\mathbf{x}^{(i)}, \mathbf{x}\right) + b \right)$$

ML

Most common Kernels

- **Polynomial Kernels**

$$K(\mathbf{x},\mathbf{z}) = (1+\mathbf{xz})^d$$

- **Radial Basis Functions**

$$K(\mathbf{x},\mathbf{z}) = \exp\frac{-(\mathbf{x}-\mathbf{z})^2}{2\sigma^2}$$

Kernels

- Provide the benefits of working in higher dimensional space
- Avoid the computational problems of working in higher dimensional space
- Avoid the theoretical curse of dimensionality problems of working in higher dimensional space

ML Support Vector Machines (SVMs)

- SVMs are Linear Learning Machines represented in a dual fashion and
- Operating in a kernel feature space

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} \phi\left(\mathbf{x}^{(i)}\right)^T \phi(\mathbf{x}) + b\right)$$

$$= \operatorname{sgn}\left(\sum_{i=1..N} \alpha_i y^{(i)} K\left(\mathbf{x}^{(i)}, \mathbf{x}\right) + b\right)$$

Other Kernel Types

- String Kernels
- Tree Kernels
- Set-based Kernels
- Etc.

ML

Maximum Margin Classifier

- SVMs control capacity by searching for maximum margin hyperplane
  - Not by reducing the number of free parameters

ML

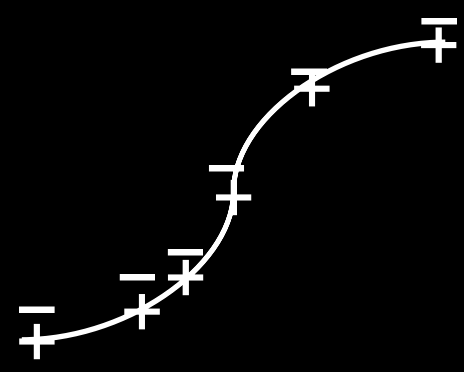
SVM Key Properties

- Duality
- Kernels
- Margin
- Convexity
- Sparseness

ML

Regression with SVMs

- Duplicate all points and add a small change
- Call the duplicates the negative class
- Solve 2-class SVM problem
- The decision boundary is the learned function



ML

Project Discussion: Evaluation

- *K*-Fold Cross Validation as means of learner eval
  - Why use CV for evaluating learning algorithms?
  - Why should you not use it?
- Should you use this for algorithm selection?

ML

Project Discussion: Evaluation

- .. *k*-fold CV eval for parameter tuning?
- .. evaluating algorithm modifications?
- What about CV tuning inside of CV eval?

Training and Test Sets I.I.D.

- Suppose your dataset is not i.i.d., instances a-b are related, c-d …, … and m-n are related
- How should you split data into training and test sets?
  – All of the data in a related group should be included in training or it should all be included in test.

Data Normalization

- **When is data normalization necessary?**
- **When is it desirable?**
- **How should you perform data normalization?**

ML

Workplans

- **Two experiments**
  - **Each with their own write-ups**
- **Plus the final paper**

ML

FLAIRS-23: Intl AI Researchers Soc.

- http://www.flairs-23.info/
- Submission deadline Nov. 23
- Daytona Beach, Florida
- May 19-21

ML FLAIRS-23: Intl AI Researchers Soc.

- http://www.flairs-23.info/ deadline: Nov 23
- Several special tracks
  - Data Mining
  - Applied Natural Language Processing
    - Book chapter
  - AI, Cognitive Sem & Comp Ling: new perspectives
  - Games and Entertainment
  - AI Planning and Scheduling
  - Learning in Intelligent Systems
  - Spatio-Temporal Reasoning
  - Uncertain Reasoning

Netflix

- Recommender systems
- Netflix
  - Round 2

ML

CMU ML Group Protests G20

- http://www.flickr.com/photos/30686429@N07/3953914015/sizes/l/in/set-72157622330082619/