MY474: Applied Machine Learning for Social Science

Lecture 8: Support Vector Machines and Active Learning

Blake Miller

27 November 2019

Agenda

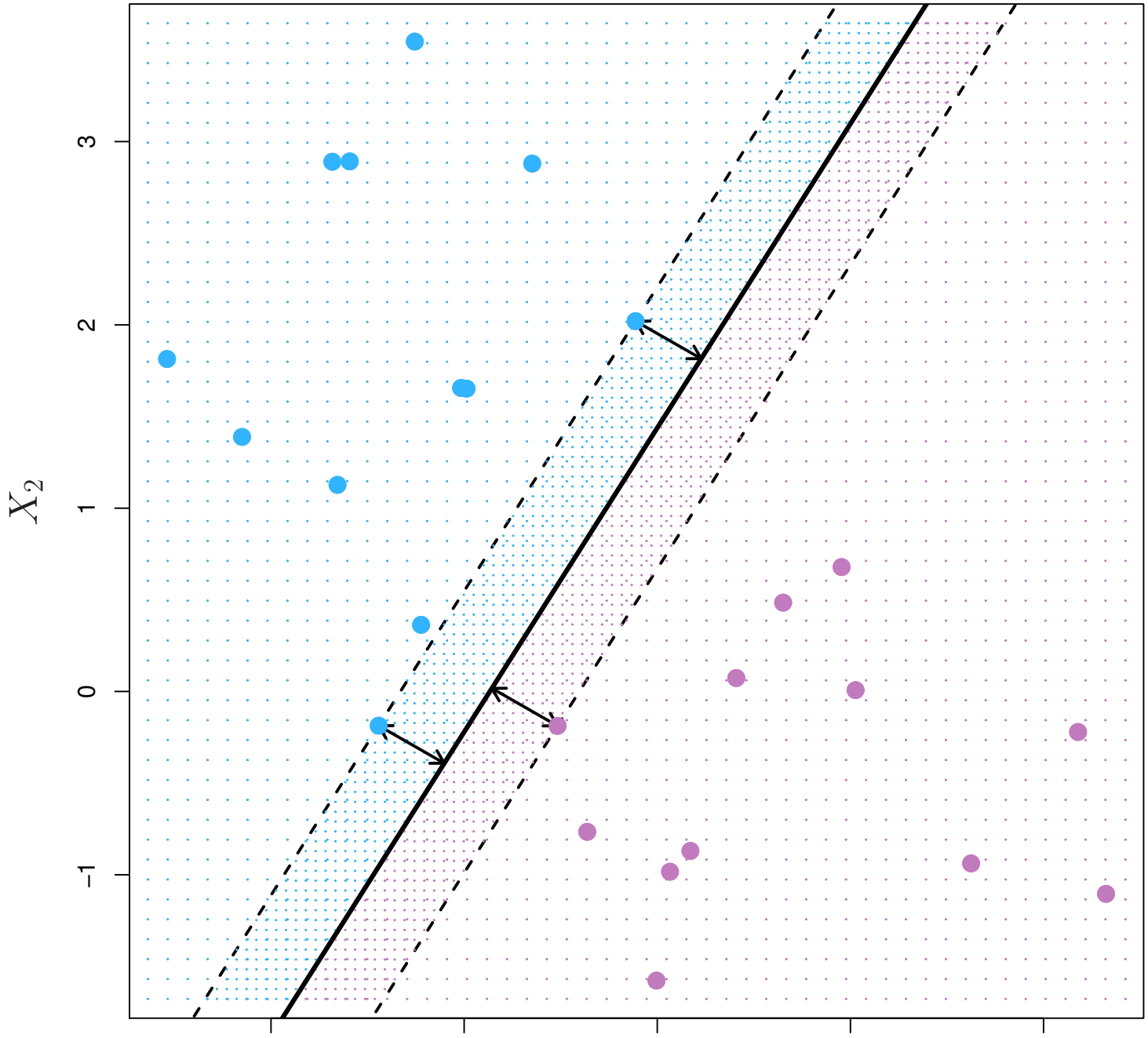
1. Main idea of SVM
2. Special case: maximal margin classifier
3. SVM
4. Comparison with logistic regression
5. Active Learning

Main idea of SVM

Support Vector Machines: main idea

- "Support-Vector Networks," Vapnik (1995)
- Consider the simplest case of two linearly separable classes (separable by a line/plane/hyperplane)
- Devise a method for finding a "good" separating hyperplane
- "Good" for SVM means with the largest margin between classes
- We will focus on 2-class problems, although there are multi-class extensions.

SVM illustration: the separable case



Hyperplanes: quick review

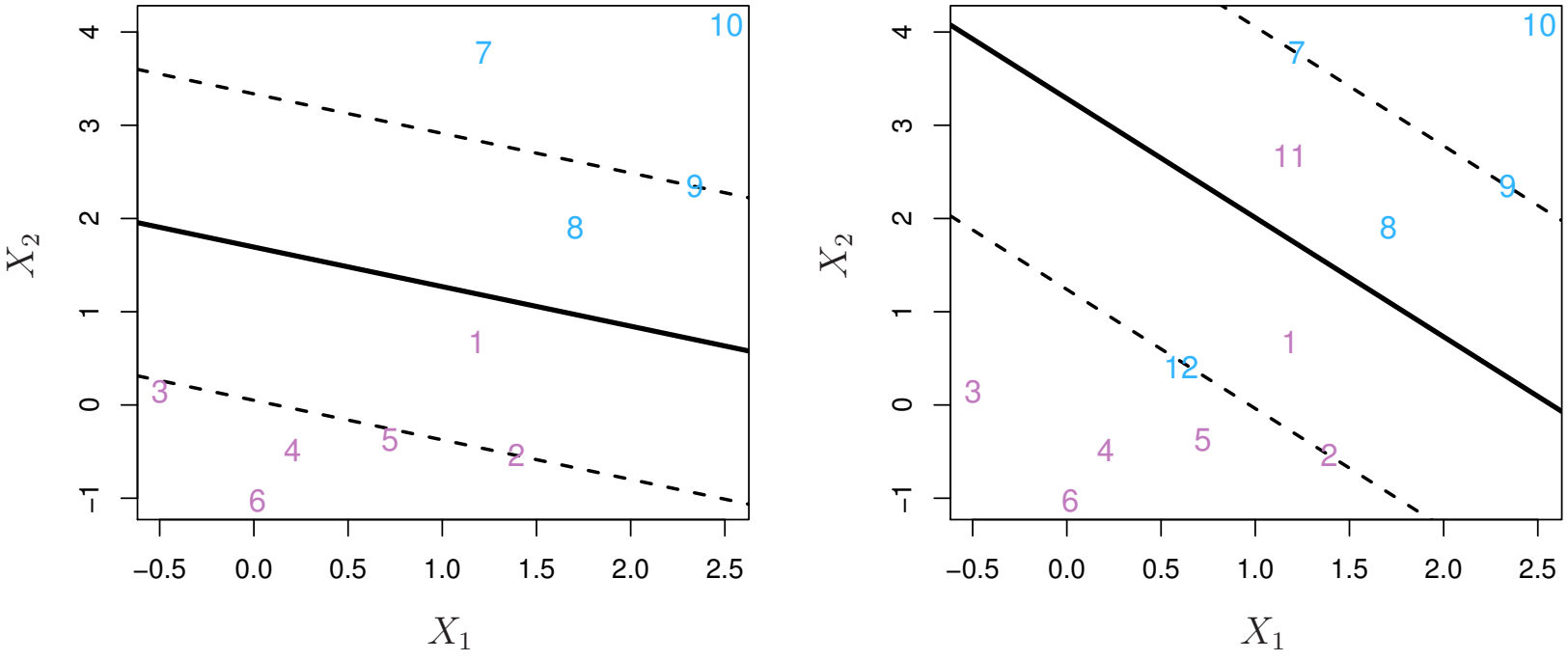
- A hyperplane in p dimensions is a flat affine subspace of dimension $p-1$.
- In general the equation for a hyperplane $F \subset R_p$ is defined by

$$F = x : \beta_0 + x'\beta = 0$$

- Equivalently: $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0$
- In $p = 2$ dimensions a hyperplane is a line.
- If $\beta_0 = 0$, the hyperplane goes through the origin, otherwise not.
- The vector $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is called the normal vector The normal vector β points in a direction orthogonal (perpendicular) to the surface of a hyperplane.
- This is because for any x_1 and x_2 in F, $(x_1 - x_2)'\beta = 0$
- Signed distance of point x to F is

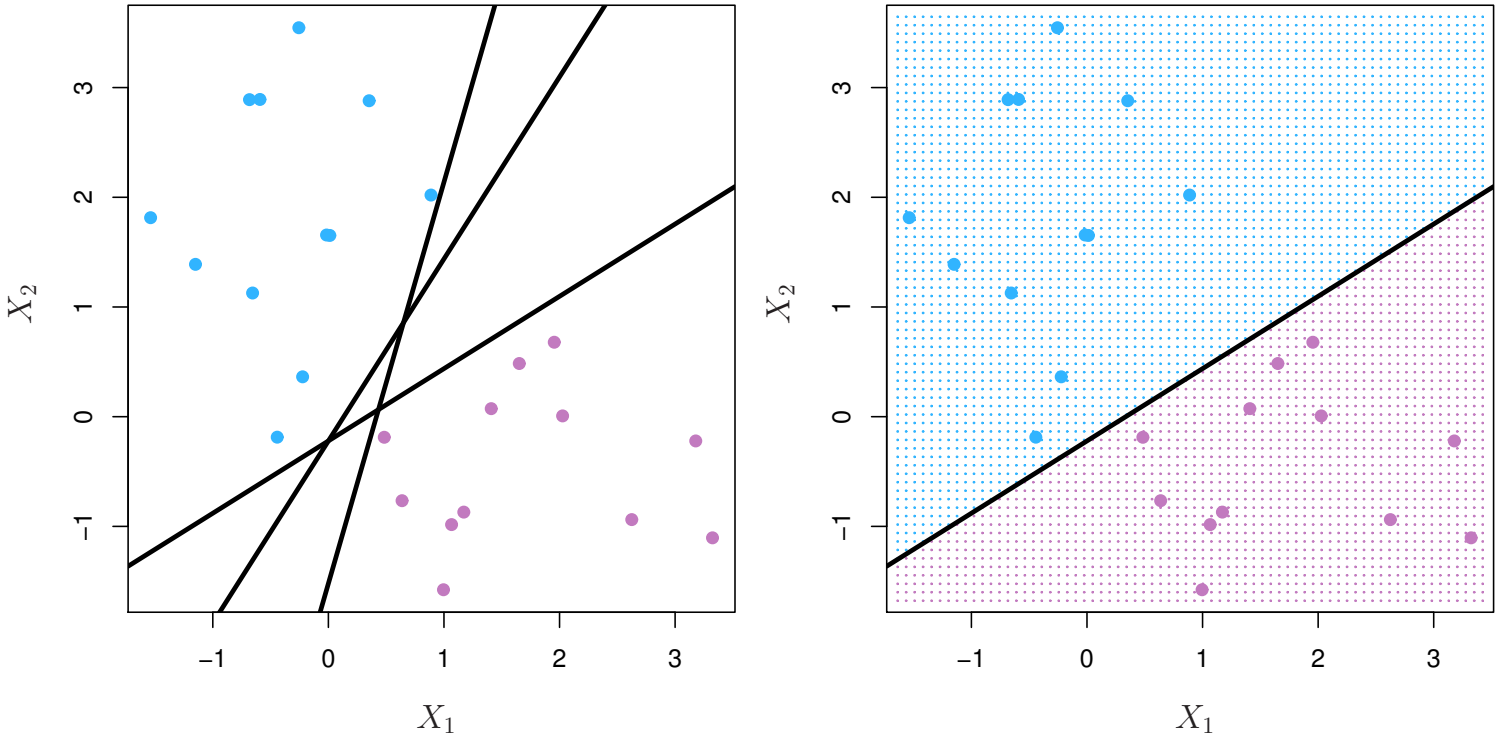
$$\frac{1}{\|\beta\|}(x'\beta + \beta_0)$$

SVM: Main concepts



- **Separating hyperplane:** a linear boundary between the classes Margin: distance from the separating hyperplane to each class (need to maximize)
- **Support vectors:** data points on the margin boundary

Separating Hyperplanes



- If $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$,then $f(X) > 0$ for points on one side of the hyperplane, and $f(X) < 0$ for points on the other.
- If we code the colored points as $Y_i = +1$ for blue, say, and $Y_i = -1$ for mauve, then if $Y_i \cdot f(X_i) > 0$ for all i, $f(X) = 0$ defines a separating hyperplane.

Classifier defined by a hyperplane

- Here we always code the class Y as $1, -1$
- Define

$$f(x) = \beta_0 + x'\beta,$$

  which is proportional to the signed distance from x to the boundary F
- For a new point x, let the classifier be

$$c(x) = \text{sign}(f(x))$$

- This type of classifier is also known as a perceptron (returning the sign of a linear combination of predictors), a term from the 50s engineering literature used in neural networks.

Maximal margin classifier

Mathematical formulation for the separable case

- For separable classes, we can always find a hyperplane such that

$$y_i(\beta_0 + x_i'\beta) > 0$$

- Want to maximize the margin, i.e. distance from the hyperplane to the nearest points
- The signed distance from x_i to $F = x : \beta_0 + x'\beta = 0$ is

$$\frac{1}{\|\beta\|}(x_i'\beta + \beta_0)$$

- $\|\beta\| = \sum_{j=1}^{p} \beta_j^2$ is the "norm" of the vector β.
- β is only determined up to a scaling factor, so introduce a constraint, $\|\beta\| = 1$
- The problem we want to solve is then

$$\max_{\beta,\beta_0} M \quad \text{subject to } \|\beta\| = 1, \quad y_i(\beta_0 + x_i'\beta) \geq M$$

Mathematical formulation for the separable case

The optimization problem

$$\max_{\beta,\beta_0} M \quad \text{subject to } \|\beta\| = 1, \quad y_i(\beta_0 + x_i'\beta) \geq M$$

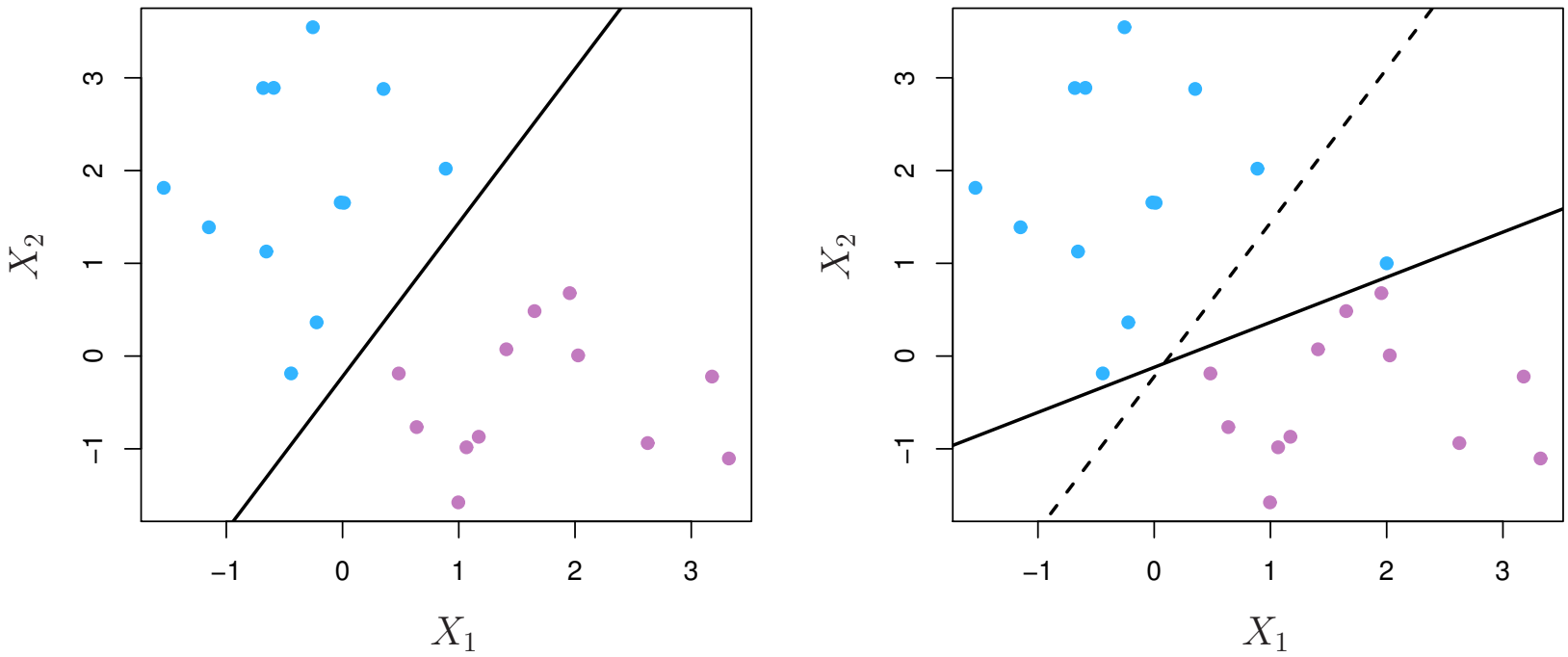
is equivalent to (Lagrange dual)

$$\min_{\beta,\beta_0} \|\beta\| \quad \text{subject to } y_i(\beta_0 + x_i'\beta) \geq 1$$

and we get that $M = 1/\|\beta\|$.

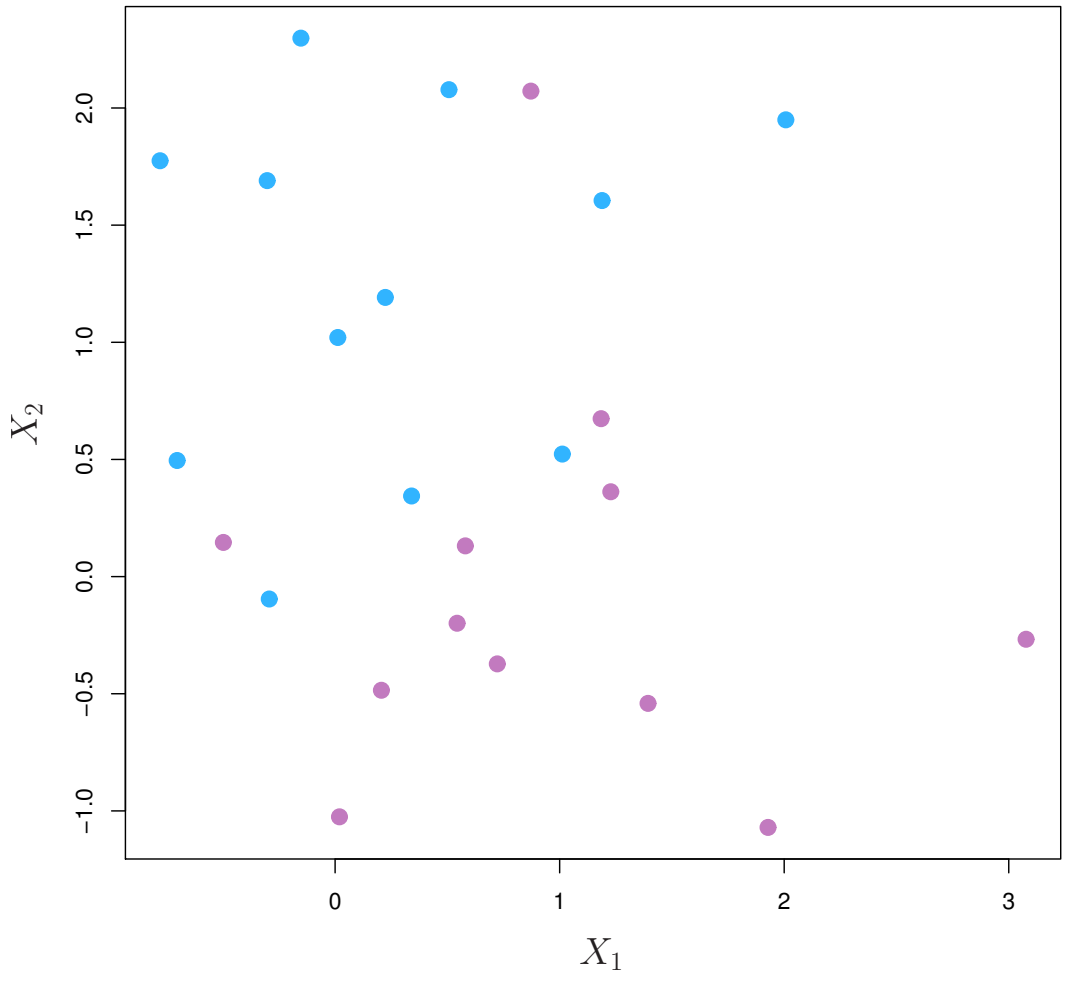
- A quadratic objective function with linear inequality constraints, hence a convex optimization problem.
- Solved using quadratic programming and Lagrange multipliers.

Problems: Noisy Data



Sometimes the data are separable, but noisy. This can lead to a poor solution for the maximal-margin classifier.

Problems: The case of non-separable classes



These data are not linearly separable meaning no separating hyperplane exists. This is often the case, unless $N < p$.

SVM

Mathematical formulation for non-separable classes

- Must allow some points on the wrong side of the support vectors Still want to maximize the margin M
- Introduce slack variables ϵ_i and replace the contraint

$$y_i(x_i'\beta + \beta_0) \geq M$$

with

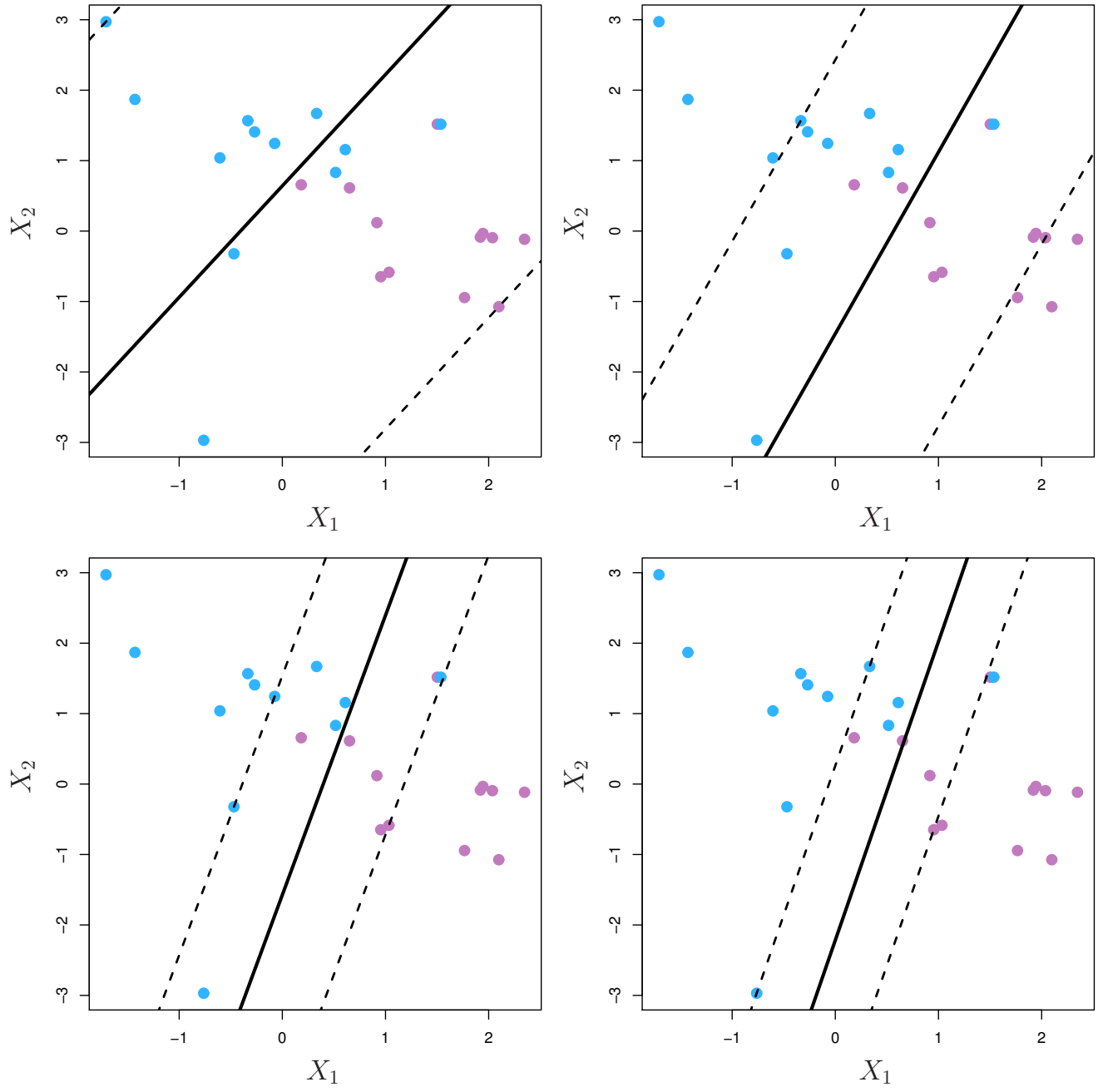
$$y_i(x_i'\beta + \beta_0) \geq M(1 - \epsilon_i)$$

- Points inside the margin but classified correctly: $0 < \epsilon_i < 1$ Misclassifications: $\epsilon_i > 1$
- Constrain the total amount of slack:

$$\epsilon_i \geq 0, \quad \sum_{i=1}^{N} \epsilon_i \leq C$$

- C is a tuning parameter

C is a regularization parameter



Mathematical formulation for non-separable classes

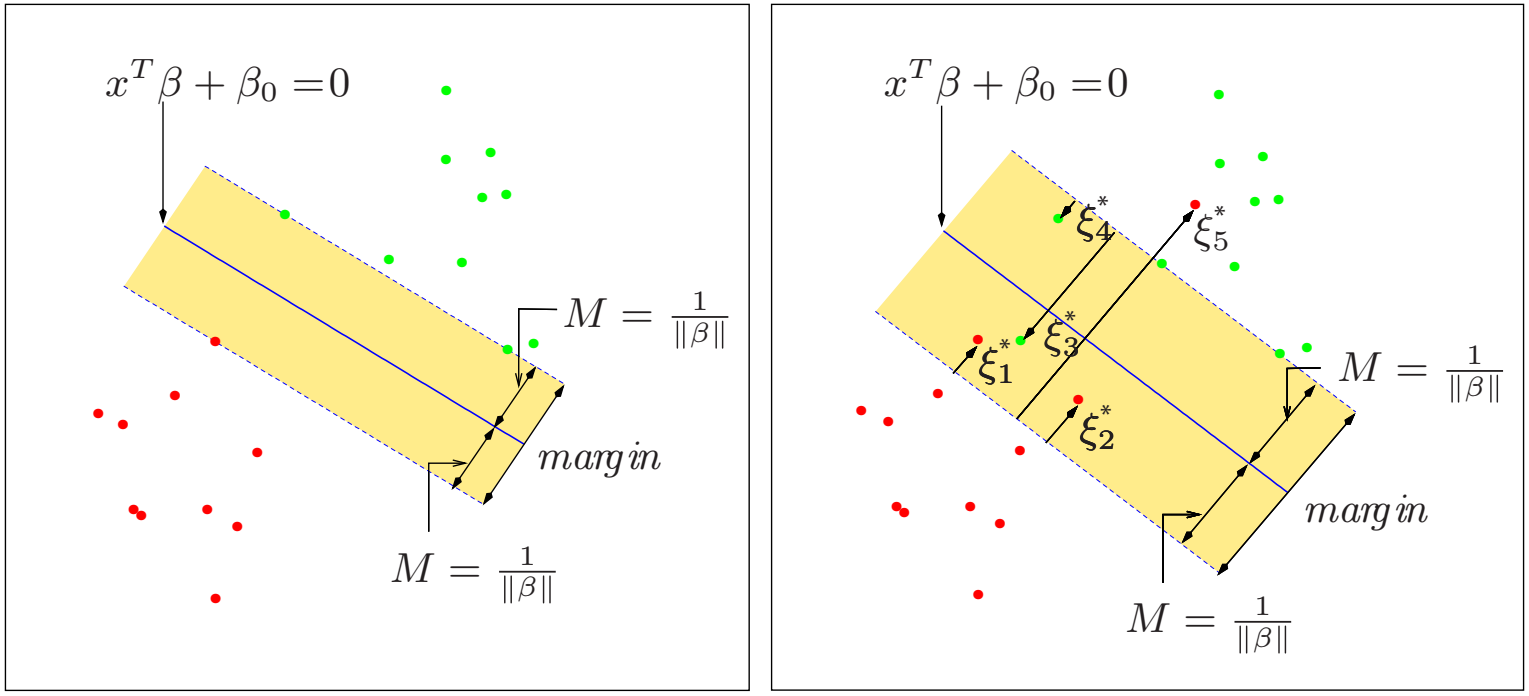
- The optimization problem is given by:

$$\max_{\beta,\beta_0} M \quad \text{subject to}$$

$$\|\beta\| = 1, \quad y_i(x_i'\beta + \beta_0) \geq M(1 - \epsilon_i), \quad \epsilon_i \geq 0, \quad \sum_{i=1}^{N} \epsilon_i \leq C$$

- Again a convex problem solved by quadratic programming with Lagrange multipliers
- Some of the ϵ_i will be exactly 0
- Support vectors are now all points that have $\epsilon_i > 0$ (points inside the margin)

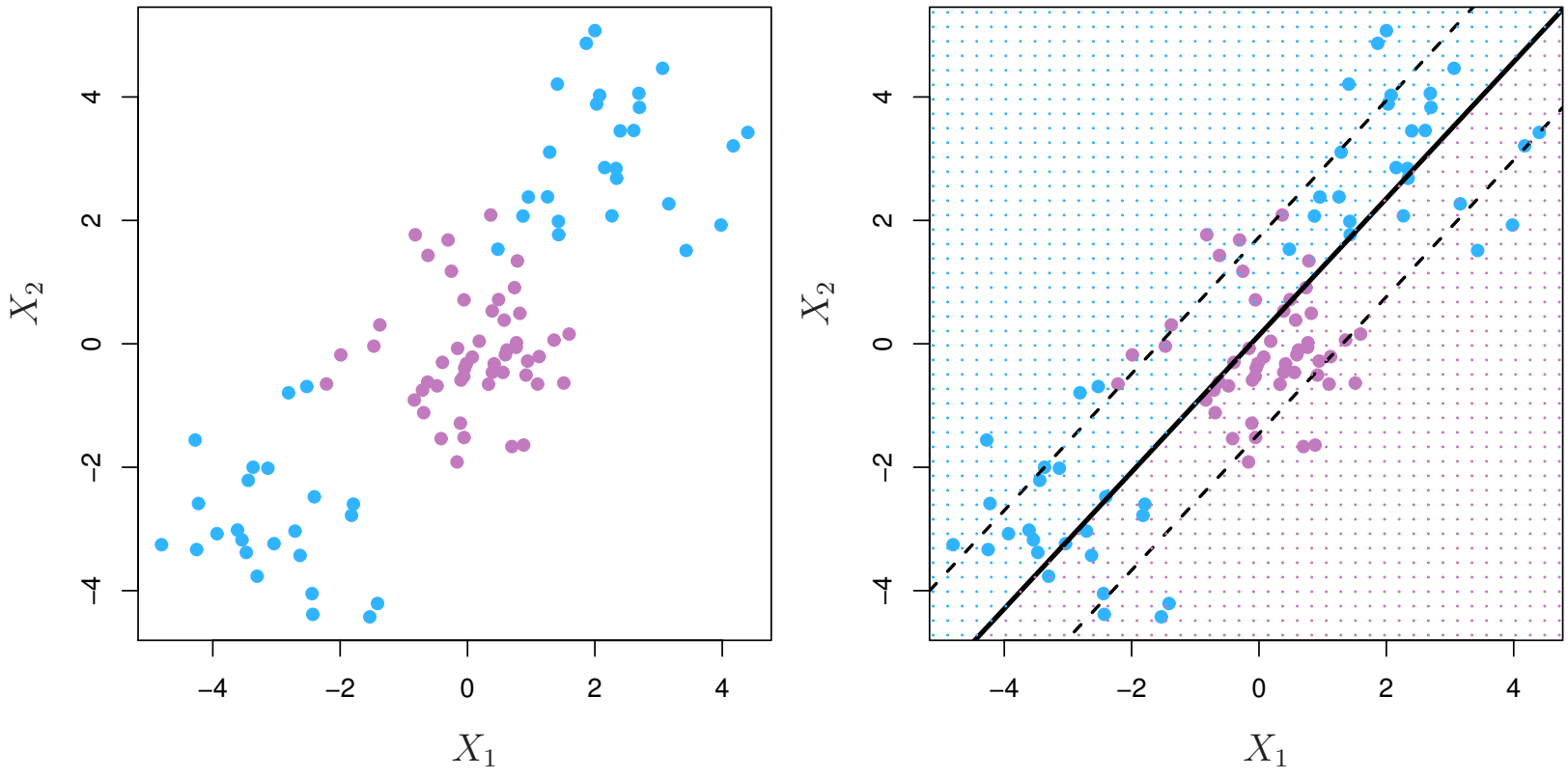
The case of non-separable classes



Left: Separable case. Right: nonseparable case (overlap) case.

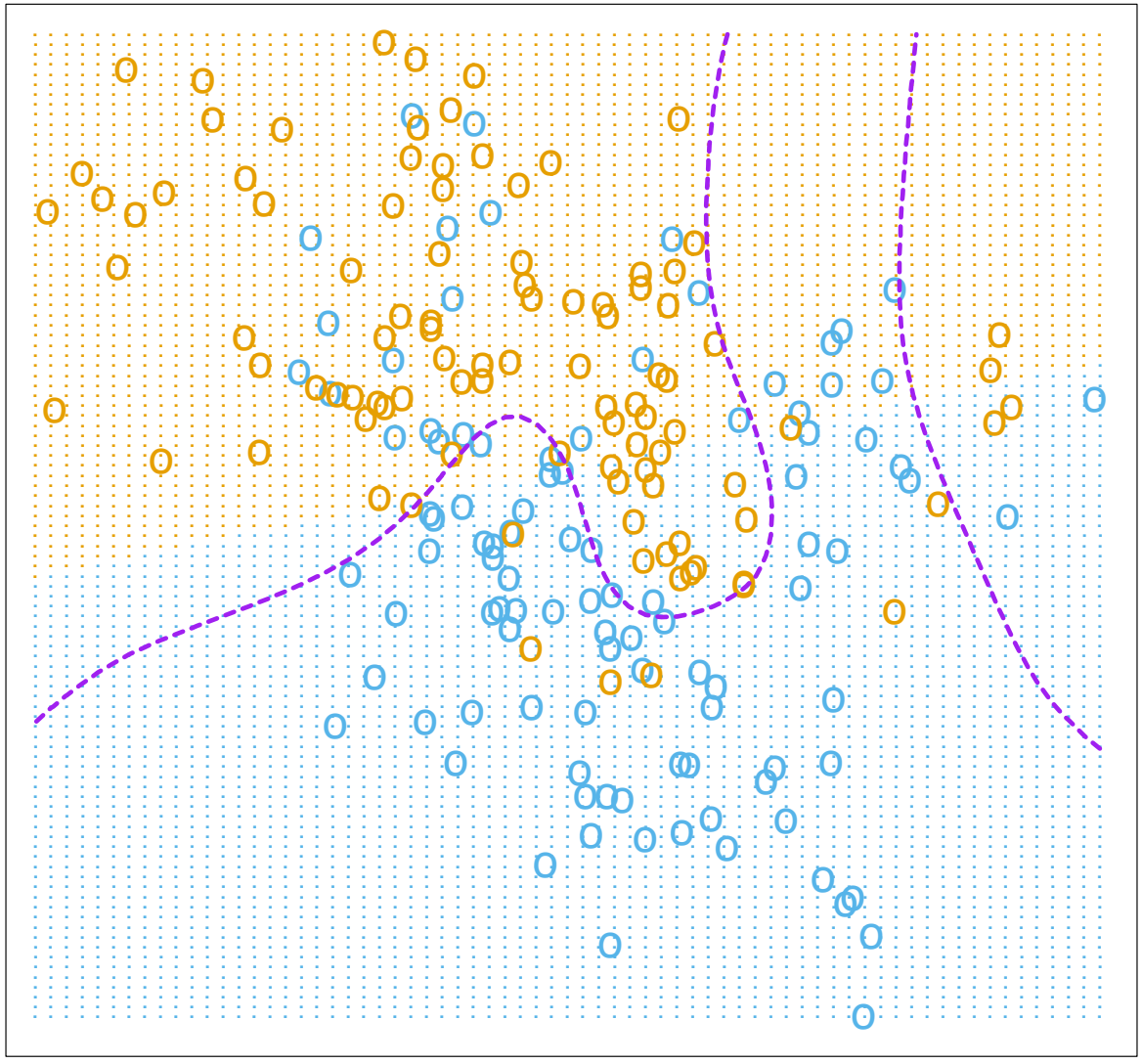
- Decision boundary is the solid line
- Broken lines bound the shaded margin of width $2M = 2/\|\beta\|$.
- The points are on the wrong side of their margin by $\epsilon_j^* = M\epsilon_j$
- Points on the correct side have $\epsilon_j^* = 0$.
- Margin is maximized subject to a total budget $\sum_{j=1}^{N} \epsilon_j^* \leq C$

Linear boundary can fail

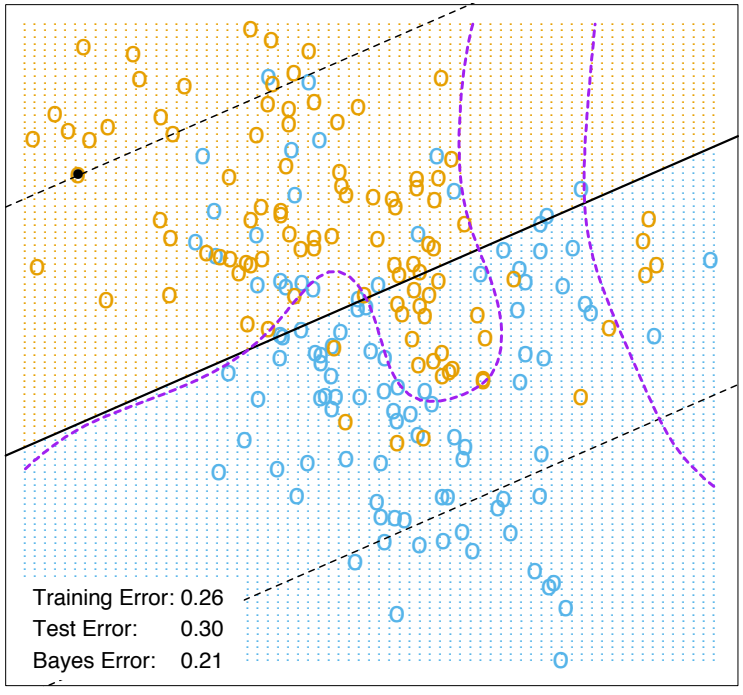


Sometime a linear boundary simply won't work, no matter what value of C. The example on the left is such a case. What to do?

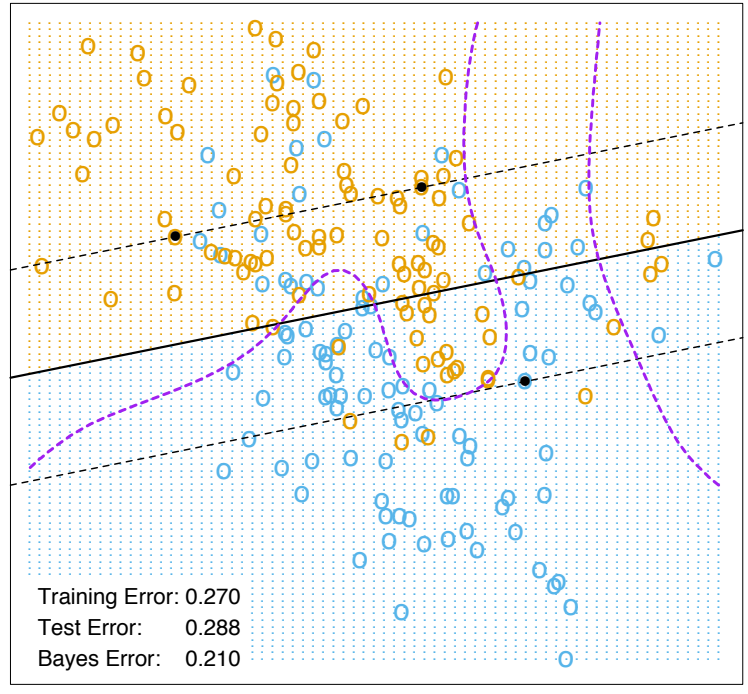
Recall the mixture example



The mixture example: Linear SVM



$C = 0.01$



$C = 10000$

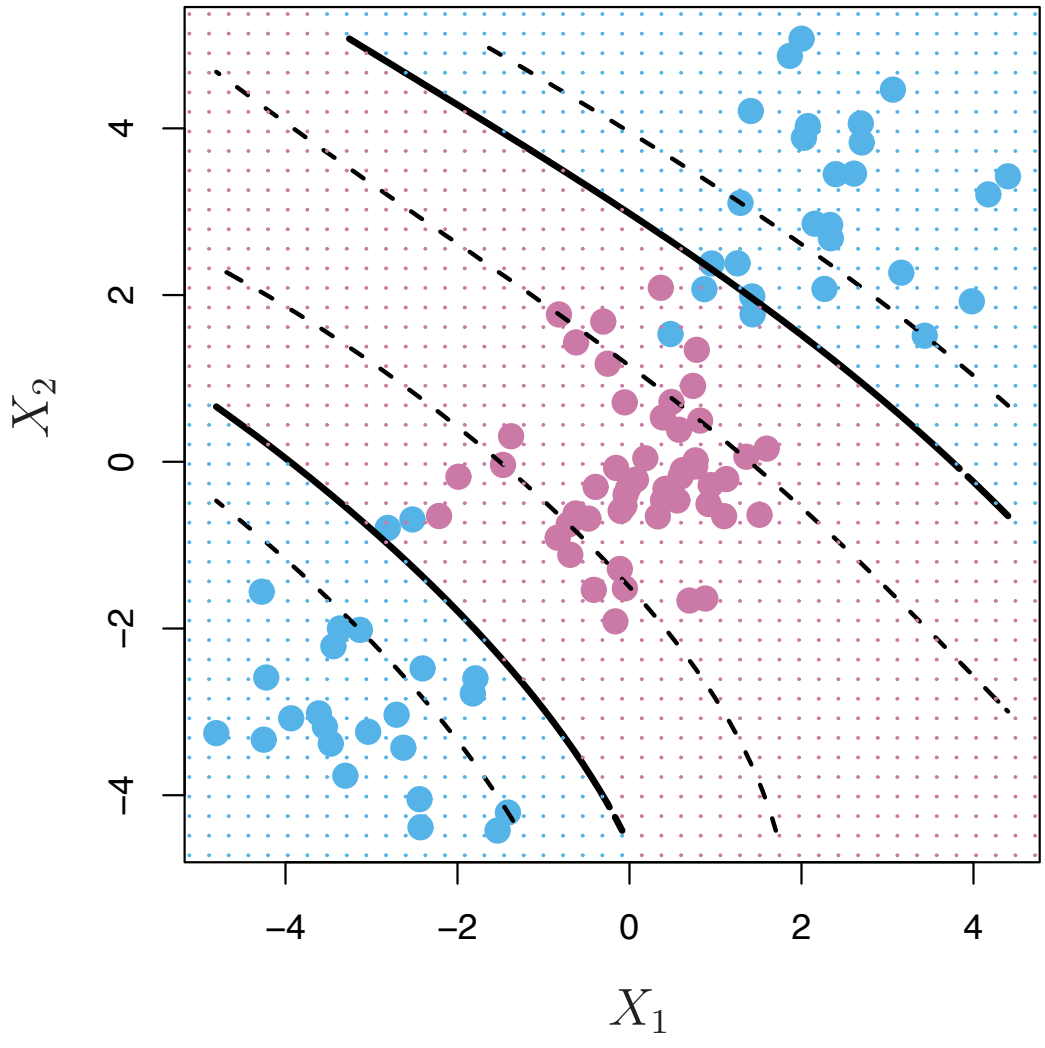
The boundary is not very sensitive to the choice of C and both perform poorly relative to the optimal decision boundary (purple)

Kernels

Feature expansion: main idea

- Embed the data in a higher-dimensional space
- For example, given p predictors $X_1, X_2, \ldots, X_p$ add new variables $X_1^2, X_2^2, \ldots, X_p^2$
- Then apply a linear method for separating the two classes
- The higher the dimension, the easier it is to find a separating hyperplane
- This results in non-linear decision boundaries in the original space.

Feature expansion example: cubic polynomial



Here we use a basis expansion of cubic polynomials to enlarge our feature space from 2 variables to 9. Our hyperplane is then defined as:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2 + \beta_5 X_1 X_2 + \beta_6 X_1^3 + \beta_7 X_2^3 + \beta_8 X_1 X_2 + \beta_9 X_1^2 X_2 = 0$$

Nonlinearities and kernels

- Polynomials (especially high-dimensional ones) get wild rather fast.
- There is a more elegant and controlled way to introduce nonlinearities in support-vector classifiers — through the use of **kernels**.
- Before we discuss these, we must understand the role of **inner products** in support-vector classifiers.

Nonlinearities and kernels (details)

- Optimization theory says the SVM solution can be written in terms of Lagrange multipliers α_i as:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x_i, x_i' \rangle$$

- In other words, to estimate the parameters $\alpha_1, \ldots, \alpha_n$ and β_0, all we need are the $\binom{n}{2} = n(n-1)/2$ inner products $\langle x_i, x_i' \rangle$ between all pairs of training observations.
- It turns out that α will only be non-zero for the set of **support points** $\mathcal{S}$
- **Support points** are the points on or within the margin
- We can add non-linearities by generalizing this inner product using kernel functions $K(x, x_i)$:

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i K(x, x_i)$$

Choices of kernels

- Linear kernel (ordinary linear SVM)

$$K(x_i, x_i') = \sum_{j=1}^{p} x_{ij} x_{i'j}$$

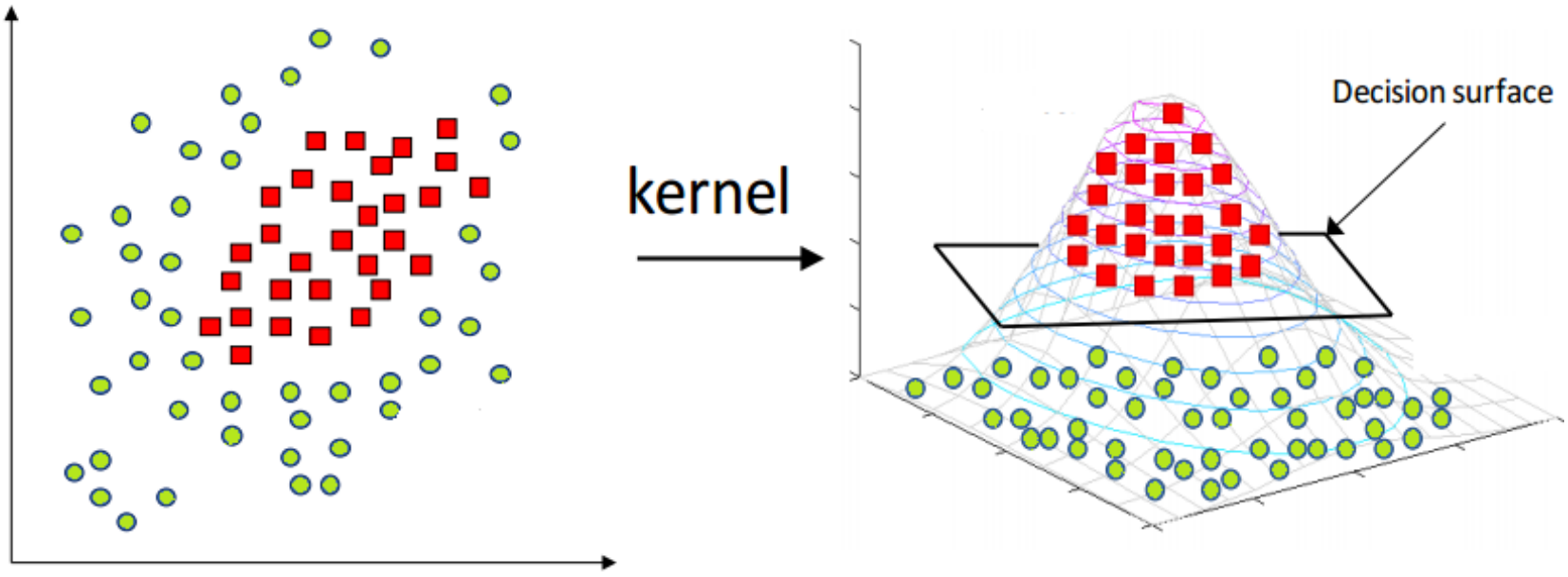
- Polynomial kernel of degree d

$$K(x_i, x_i') = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d$$

- Radial basis kernel of degree d with hyperparameter γ

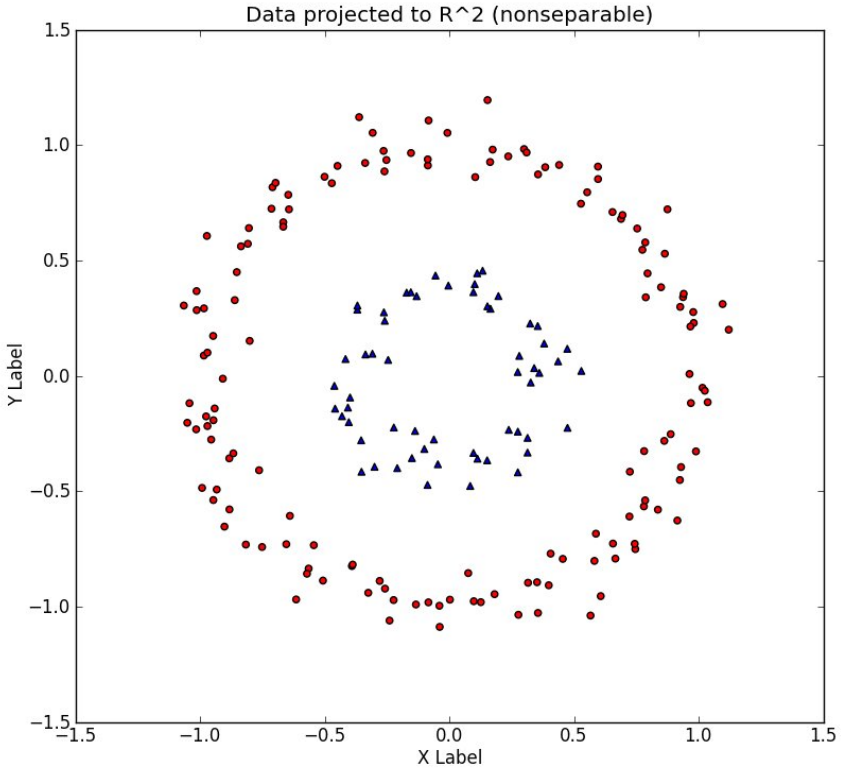
$$K(x_i, x_i') = exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$$

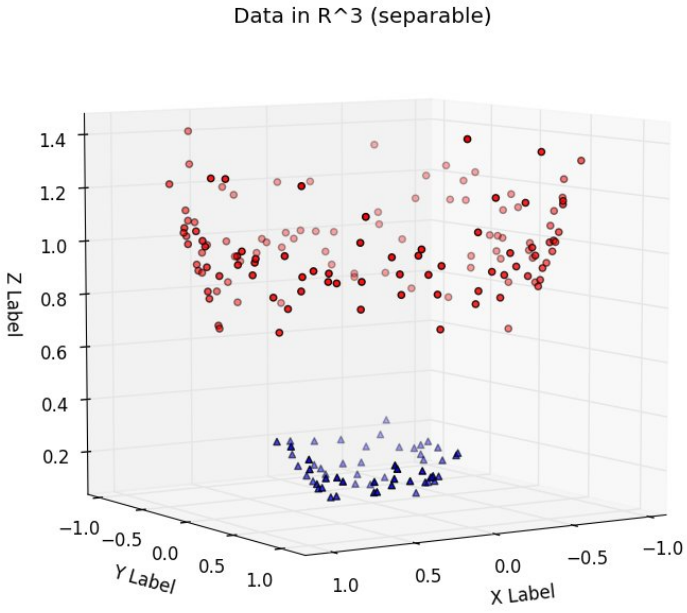
Visualizing kernel SVM (Donut data)



The polynomial kernel applied to linearly non-separable data.

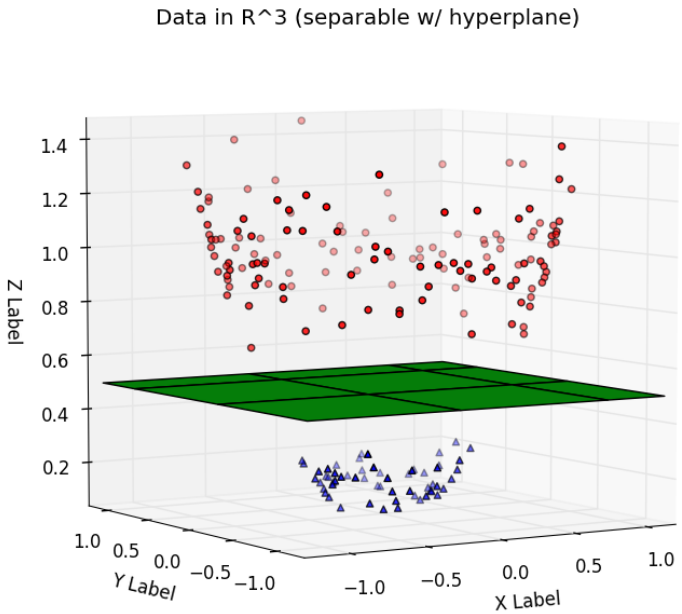
Visualizing kernel SVM (Donut data)

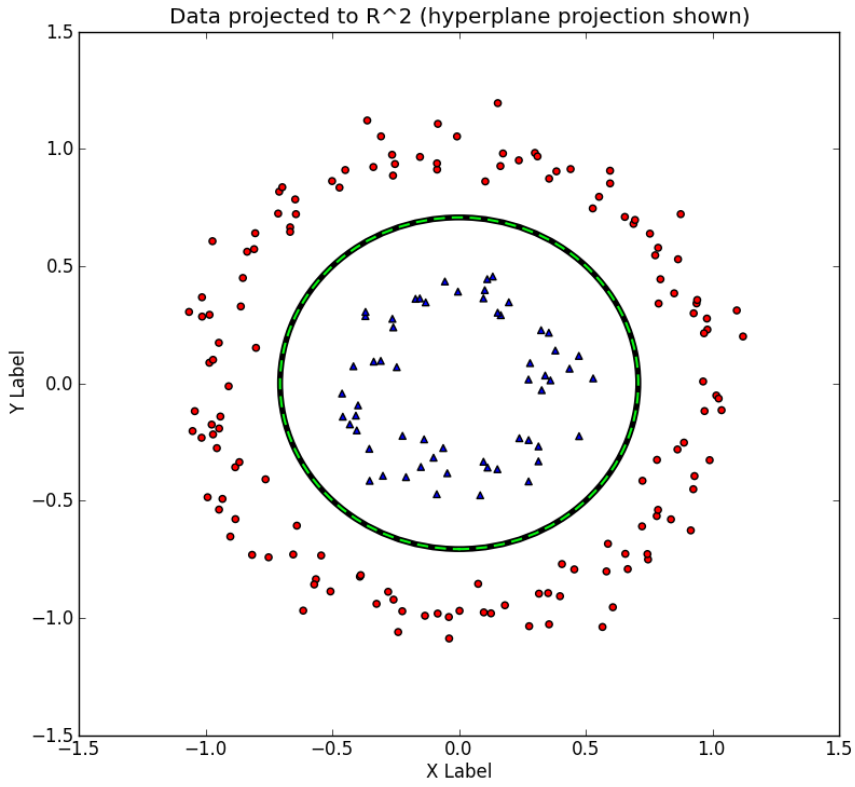




The polynomial kernel applied to linearly non-separable data.

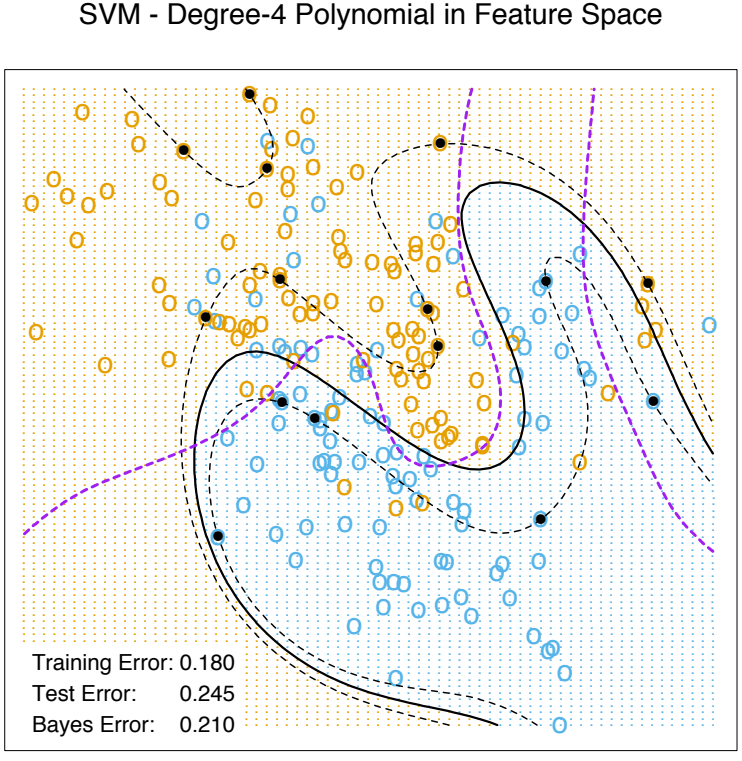
Visualizing kernel SVM (Donut data)

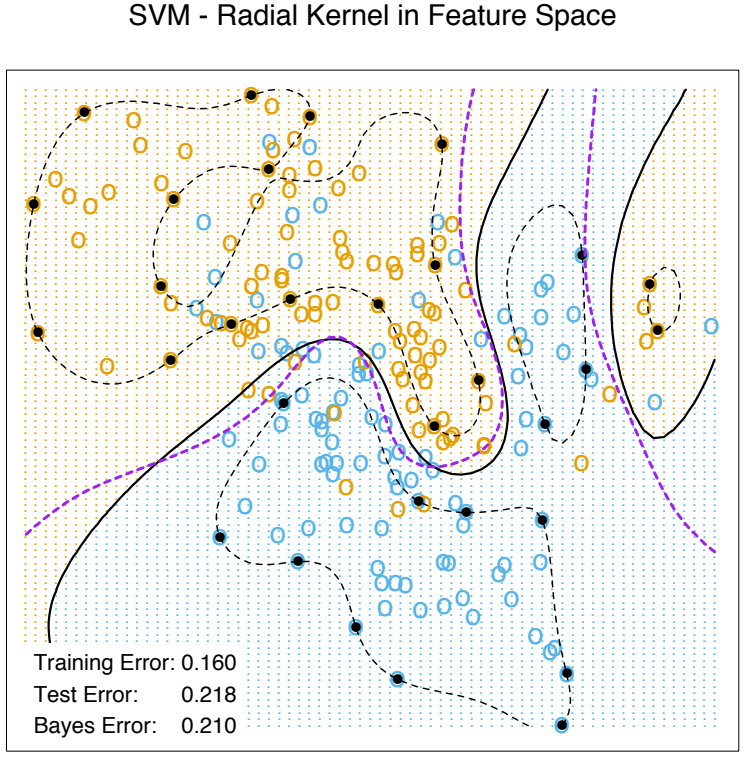




The polynomial kernel applied to linearly non-separable data.

The mixture example: Kernel SVM





Polynomial and radial kernels perform quite well compared to the bayes optimal classifier (purple)

SVMs: more than 2 classes?

- The SVM as defined works for $K = 2$ classes. What do we do if we have $K > 2$ classes?
  - **OVA** (One versus All): Fit K different 2-class SVM classifiers $\hat{f}_k(x)$, $k = 1, \ldots, K$; each class versus the rest. Classify x^* to the class for which $\hat{f}_k(x^*)$ is largest.
  - **OVO** (One versus One). Fit all Kpairwise classifiers $\hat{f}_{k\ell}(x)$. Classify x^* to the class that wins the most $k\ell$ pairwise competitions.
- Which to choose? If K is not too large, use OVO.

Comparison with logistic regression

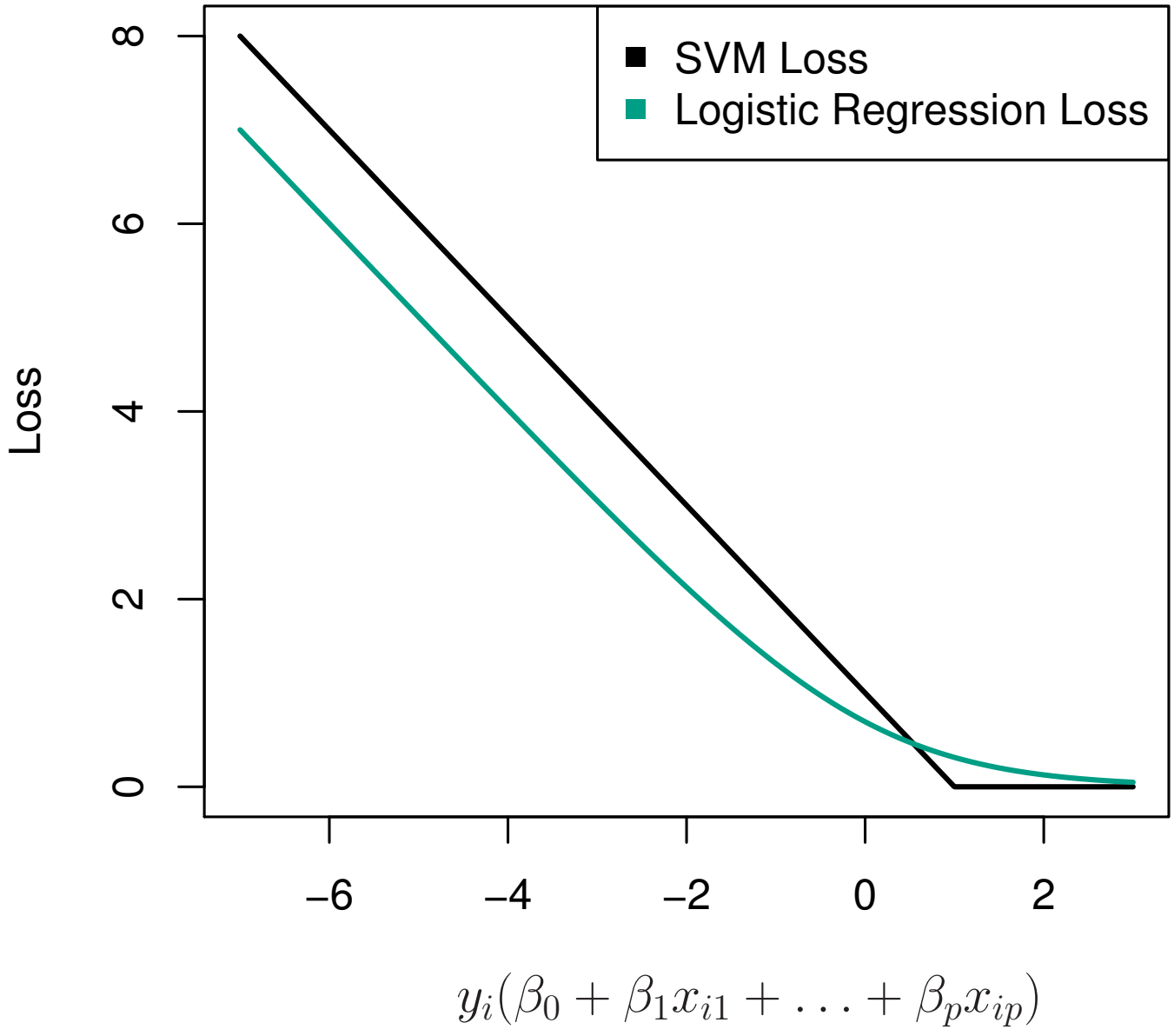
SVM versus logistic regression

- With $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, we can rephrase support-vector classifier optimization as

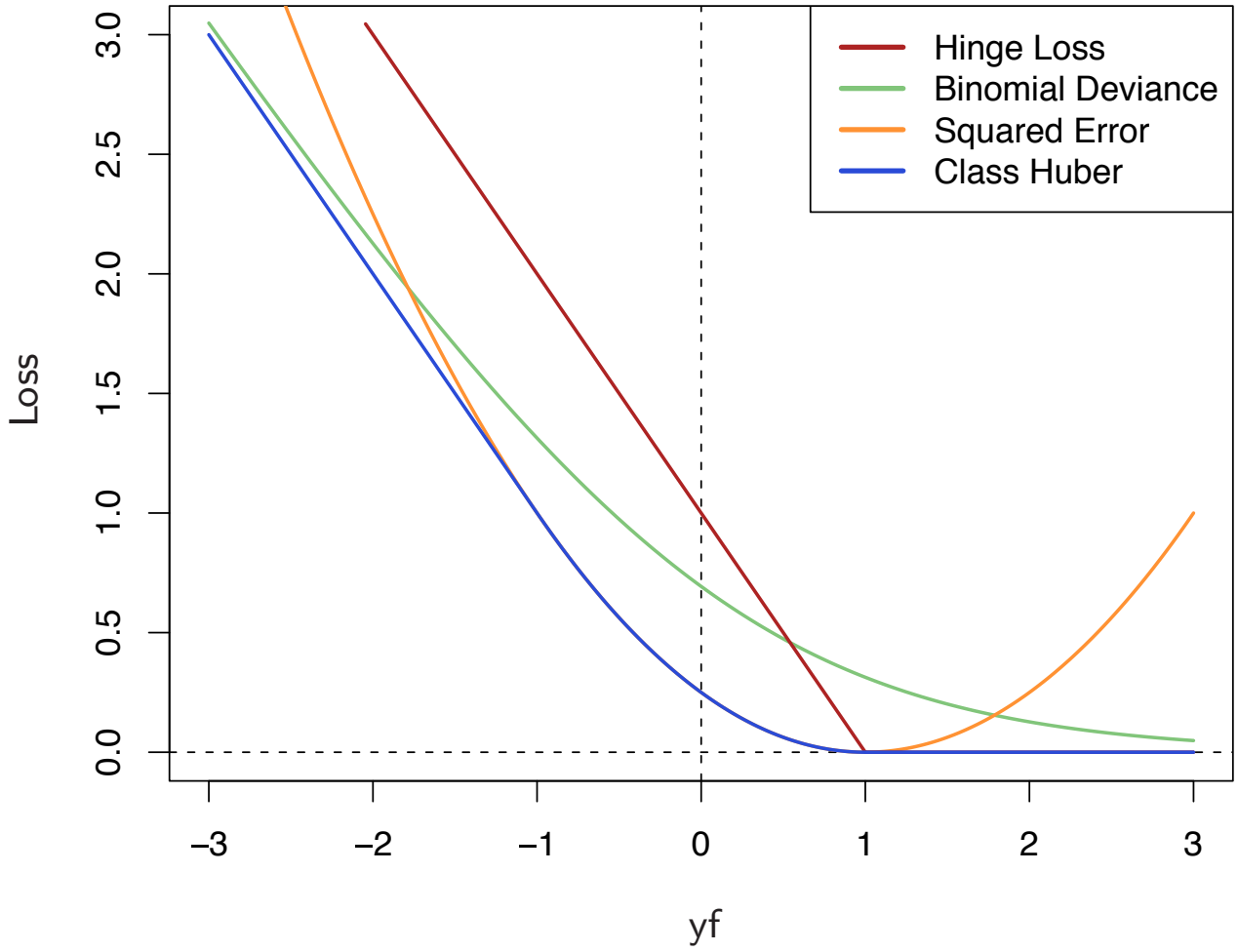
$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \left\{ \sum_{i=1}^{n} max[0, 1 - y_i f(x_i)] + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

- Logistic regression replaces the hinge loss $(1 - yf)+$ with the negative log-likelihood $ln(1 + e^{-yf})$.
- The hinge loss is very similar to the negative log-likelihood in logistic regression (see next slide).

SVM versus logistic regression: loss function



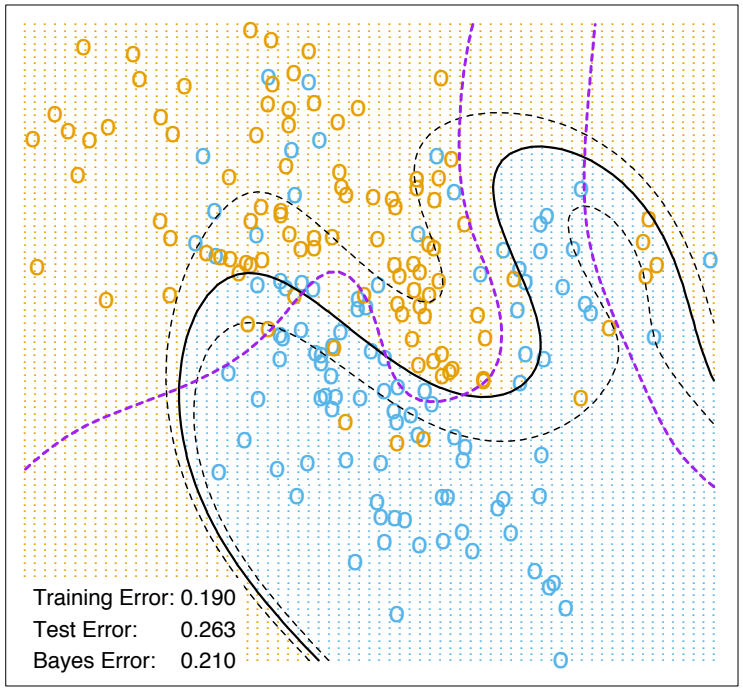
There are many loss functions!



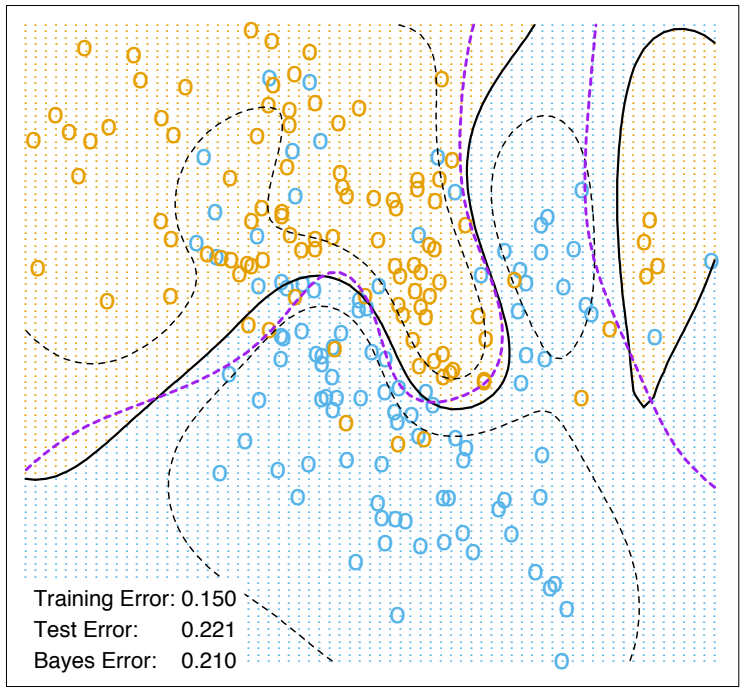
Don't worry, you don't have to know these. Just an FYI.

Support Vector versus Logistic Regression

LR - Degree-4 Polynomial in Feature Space



LR - Radial Kernel in Feature Space



Kernel logistic regression gives very similar results to kernel SVM.

Support Vector versus Logistic Regression

- LR classification performance is usually similar to the SVM.
- When classes are (nearly) separable, SVM does better than LR.
- When not, LR (with ridge penalty) and SVM very similar.
- LR provides estimates of class probabilities.
- LR naturally generalizes to the multi-class case (instead of relying on OVO, OVA).
- For nonlinear boundaries, kernel SVMs are popular. Can use kernels with LR as well, but computations are more expensive.

SVM application: text classification

- Why might kernel SVM not perform well with text data?
  - Most text classification problems are linearly separable.
  - Text data has many features so increasing the dimensionality does not usually improve performance.
  - Even in rare cases where a kernel marginally improves performance, it may not be worth it due to the computational cost.
  - With kernel SVM, there are more hyperparameters to tune!

Active Learning

A Motivating Example: Identifying Govt. Weibo Accounts

Weibo account of the Xinjiang Provincial Public Security Bureau

A Motivating Example: Identifying Govt. Weibo Accounts

- **Research Problem:**

A Motivating Example: Identifying Govt. Weibo Accounts

- **Research Problem:**
  - Examine social media as an **input institution** in authoritarian politics.

A Motivating Example: Identifying Govt. Weibo Accounts

- **Research Problem:**
  - Examine social media as an **input institution** in authoritarian politics.
  - Identify **astroturfers** affiliated with these accounts and examine the logics of **government astroturfing campaigns.**

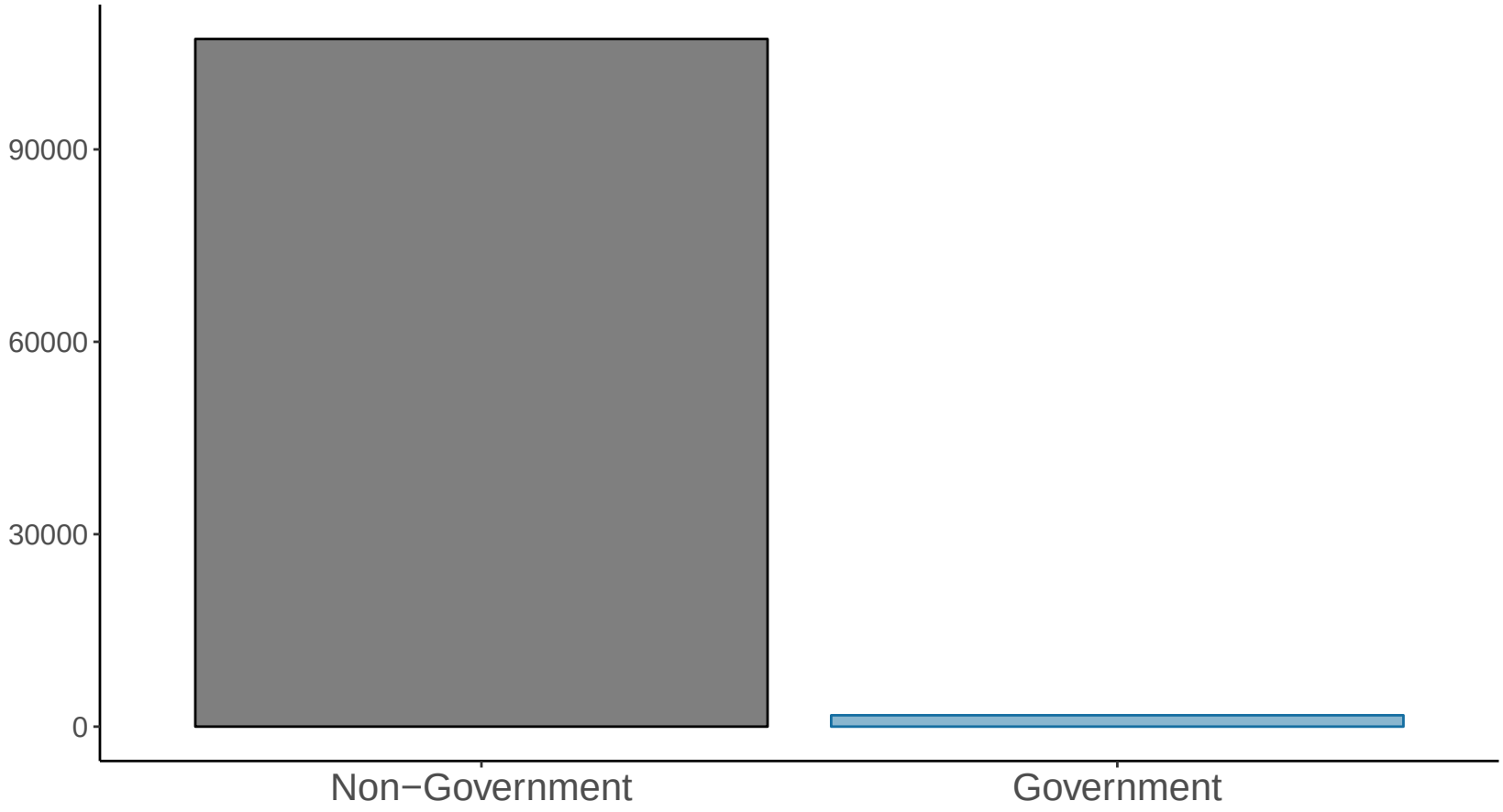
A Motivating Example: Identifying Govt. Weibo Accounts

- **Research Problem:**
  - Examine social media as an **input institution** in authoritarian politics.
  - Identify **astroturfers** affiliated with these accounts and examine the logics of **government astroturfing campaigns.**
- **Data Collection Goal:** Retrieve all government accounts from a large sample of **Weibo** accounts.

A Motivating Example: Identifying Govt. Weibo Accounts

- **Research Problem:**
  - Examine social media as an **input institution** in authoritarian politics.
  - Identify **astroturfers** affiliated with these accounts and examine the logics of **government astroturfing campaigns.**
- **Data Collection Goal:** Retrieve all government accounts from a large sample of **Weibo** accounts.
- **Method:** Manually annotate a random sample of accounts; train a text classifier on the account name and description

A Motivating Example: Identifying Govt. Weibo Accounts



One major concern: the two classes for this problem are highly imbalanced.

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts
  - Need to label 1 million accounts

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts
  - Need to label 1 million accounts
- **Assume:** Research assistants can code 10 accounts every minute on average.

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts
  - Need to label 1 million accounts
- **Assume:** Research assistants can code 10 accounts every minute on average.
- **Time:** Just under 10 months of M-F 9am to 5pm to achieve this goal.

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts
  - Need to label 1 million accounts
- **Assume:** Research assistants can code 10 accounts every minute on average.
- **Time:** Just under 10 months of M-F 9am to 5pm to achieve this goal.
- **Labor Cost:** National living wage (£7.83/hour), total labor cost of ~£13,000

A Motivating Example: Identifying Govt. Weibo Accounts

- **Problem:** Government accounts are quite rare
  - $\approx$ 1 in every 2,000 accounts
- **Goal:** sample 500 government accounts
  - Need to label 1 million accounts
- **Assume:** Research assistants can code 10 accounts every minute on average.
- **Time:** Just under 10 months of M-F 9am to 5pm to achieve this goal.
- **Labor Cost:** National living wage (£7.83/hour), total labor cost of ~£13,000
- **Management Cost:** We would also have to spend time managing and checking reliability of the RA's work

A Motivating Example: Identifying Govt. Weibo Accounts

With only 500 government accounts in our training data, our model will likely **perform quite poorly.**

Active vs. Passive Learning

Weibo account of the Xinjiang Provincial Public Security Bureau

Active vs. Passive Learning

Weibo account of the Xinjiang Provincial Public Security Bureau

Astroturfer Detection

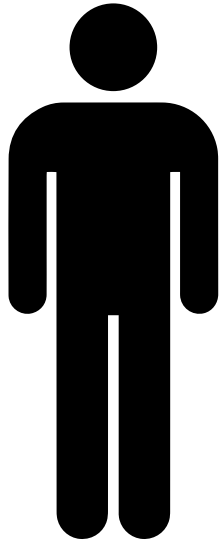
Are Weibo accounts
government or **non-government**?

	Party	government	good	football	China	…
X_1	8	9	5	0	12	…
X_2	0	0	8	4	0	…
X_3	0	0	7	4	2	…
X_4	9	8	6	0	8	…
…	…	…	…	…	…	…

Passive Learning

	Party	government	good	football	China	...
X_1	8	9	5	0	12	...
X_2	0	0	8	4	0	...
X_3	0	0	7	4	2	...
X_4	9	8	6	0	8	...
...	...	...	...	...	...	...

Random Sample



Expert/Oracle

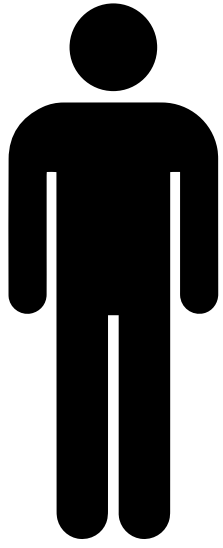
Passive Learning

	Party	government	good	football	China	...
x_1	8	9	5	0	12	...
x_2	0	0	8	4	0	...
x_3	0	0	7	4	2	...
x_4	9	8	6	0	8	...
...	...	...	...	...	...	...

Random Sample

x	y
x_1	**GOV**
x_2	**¬GOV**
x_3	**¬GOV**
x_4	**GOV**
...	...

An expert labels each account as **government** or **non-government**

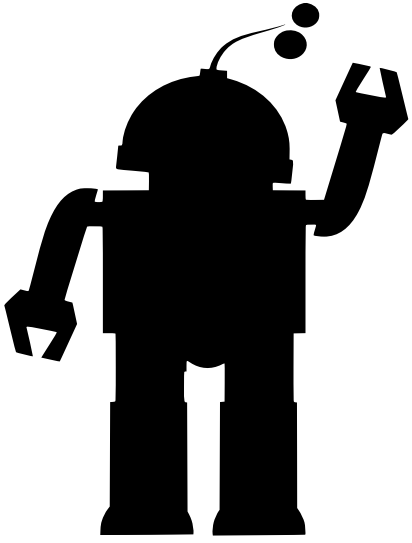


Expert/Oracle

Passive Learning

	Party	government	good	football	China	...
x_1	8	9	5	0	12	...
x_2	0	0	8	4	0	...
x_3	0	0	7	4	2	...
x_4	9	8	6	0	8	...
...	...	...	...	...	...	...

Random Sample

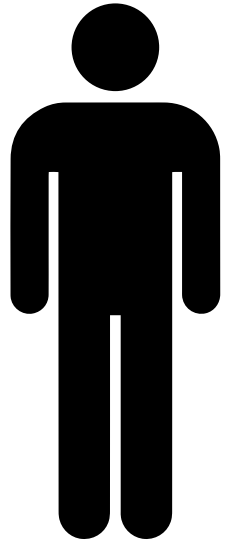


Classifier

Train a classifier on labeled documents

x	y
x_1	GOV
x_2	¬GOV
x_3	¬GOV
x_4	GOV
...	...

An expert labels each account as **government** or **non-government**



Expert/Oracle

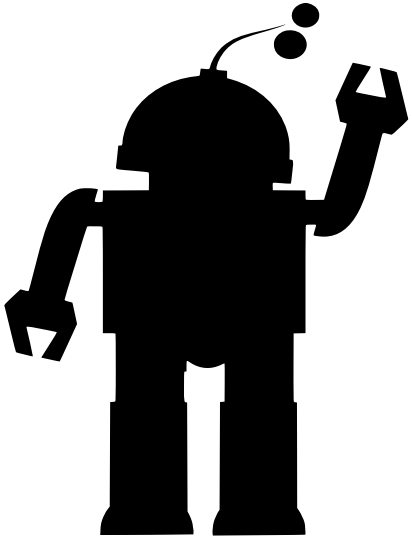
Passive Learning

	Party	government	good	football	China	...
x_1	8	9	5	0	12	...
x_2	0	0	8	4	0	...
x_3	0	0	7	4	2	...
x_4	9	8	6	0	8	...
...	...	...	...	...	...	...

x^*	12	2	7	0	2	...

Random Sample

Input unlabeled document

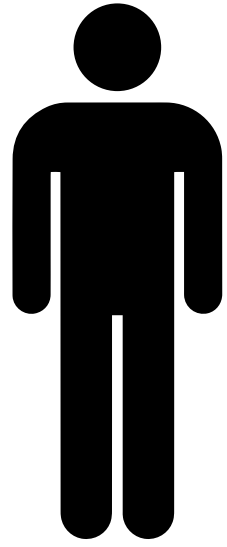


Classifier

Train a classifier on labeled documents

x	y
x_1	GOV
x_2	¬GOV
x_3	¬GOV
x_4	GOV
...	...

An expert labels each account as **government** or **non-government**



Expert/Oracle

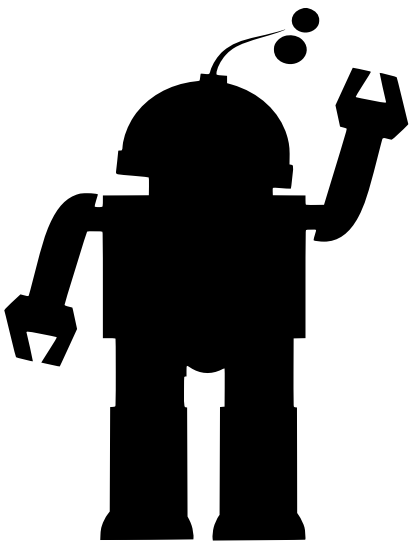
Passive Learning

	Party	government	good	football	China	...
x_1	8	9	5	0	12	...
x_2	0	0	8	4	0	...
x_3	0	0	7	4	2	...
x_4	9	8	6	0	8	...
...	...	...	...	...	...	...

x^*	12	2	7	0	2	...

Random Sample

Input unlabeled document



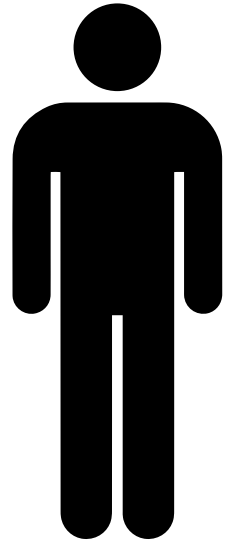
Classifier

government

x	y
x_1	GOV
x_2	¬GOV
x_3	¬GOV
x_4	GOV
...	...

Train a classifier on labeled documents

An expert labels each account as **government** or **non-government**

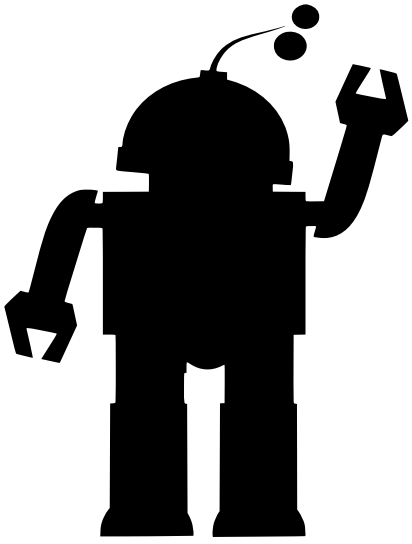


Expert/Oracle

Active Learning

	Party	government	good	football	China	...
X_1	8	9	5	0	12	...
X_2	0	0	8	4	0	...
X_3	0	0	7	4	2	...
X_4	9	8	6	0	8	...
...	...	...	...	...	...	...

The classifier examines unlabeled documents

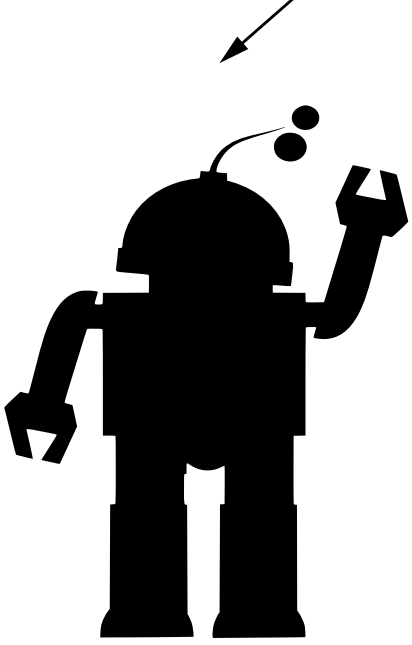


Classifier

Active Learning

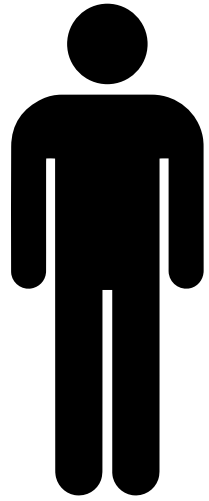
	Party	government	good	football	China	...
X_1	8	9	5	0	12	...
X_2	0	0	8	4	0	...
X_3	0	0	7	4	2	...
X_4	9	8	6	0	8	...
...	...	...	...	...	...	...

The classifier examines unlabeled documents



x*?

The classifier asks the expert to label a document it has trouble classifying



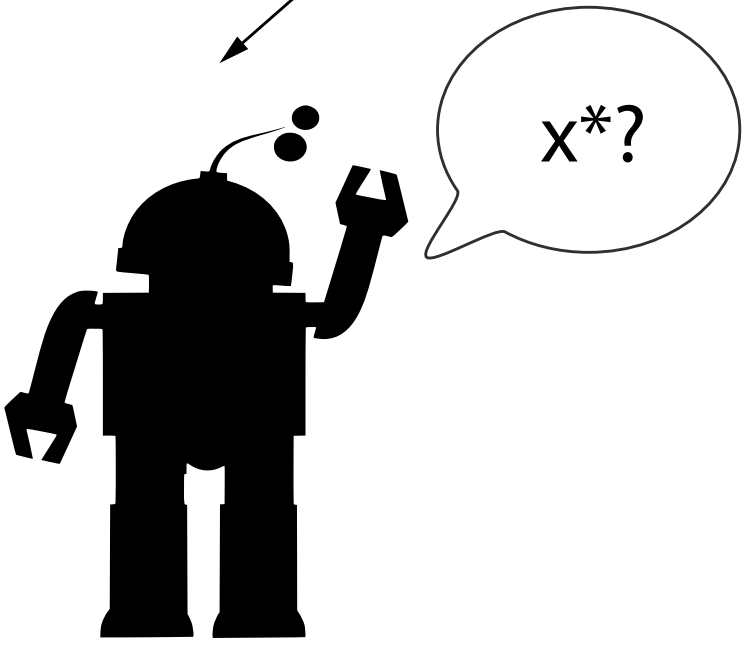
Classifier

Expert/Oracle

Active Learning

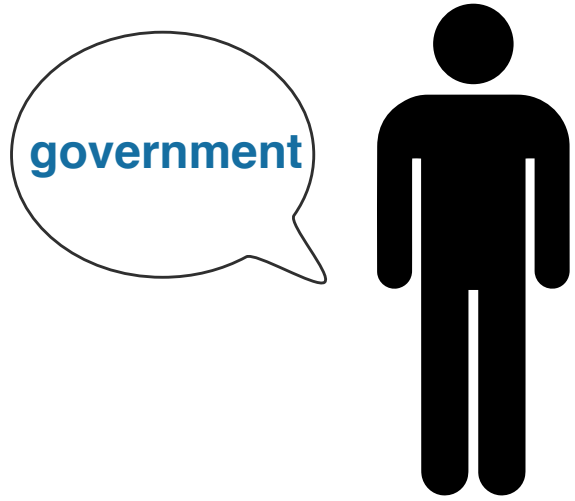
	Party	government	good	football	China	...
X_1	8	9	5	0	12	...
X_2	0	0	8	4	0	...
X_3	0	0	7	4	2	...
X_4	9	8	6	0	8	...
...	...	...	...	...	...	...

The classifier examines unlabeled documents



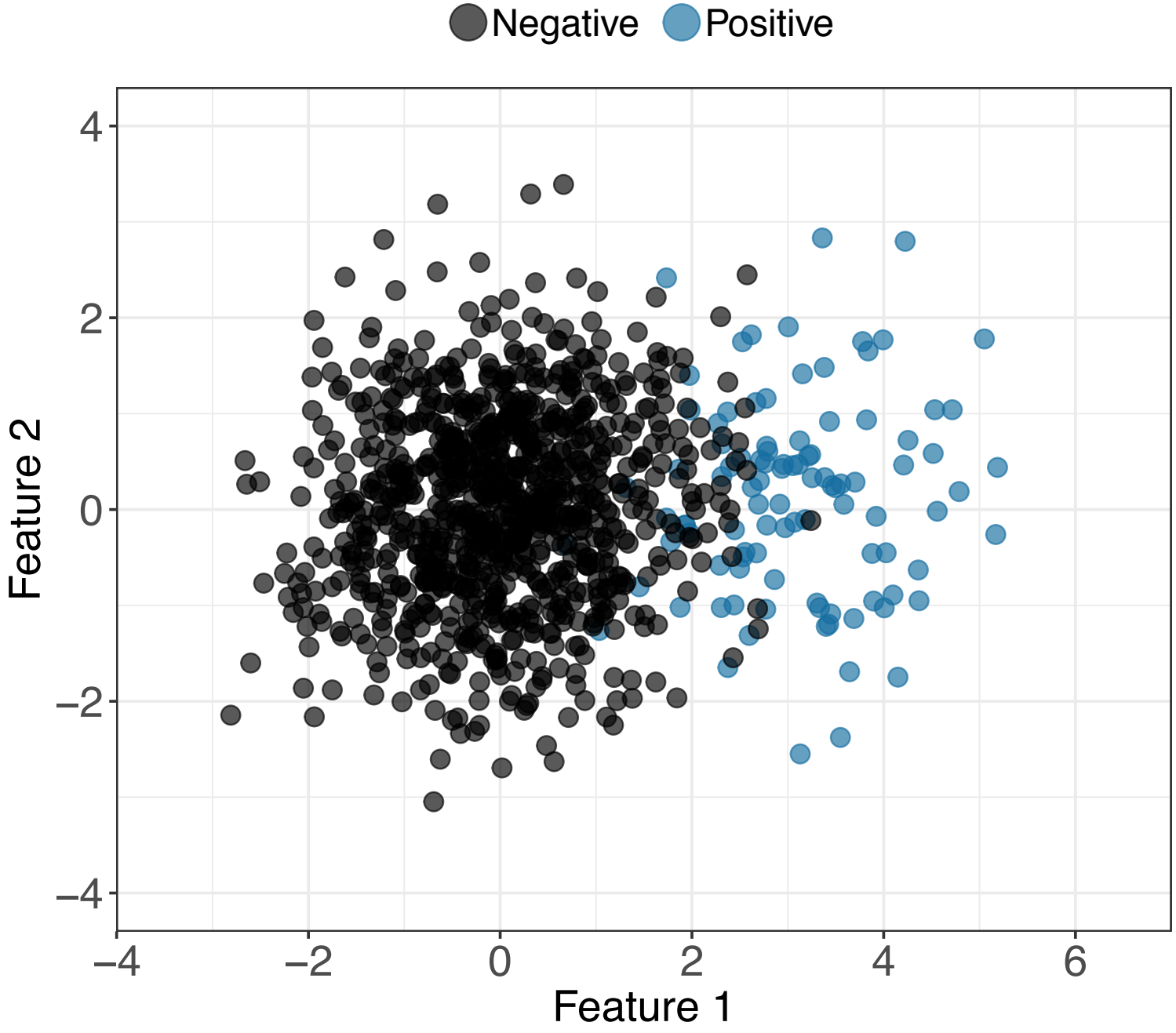
Classifier

The classifier asks the expert to label a document it has trouble classifying

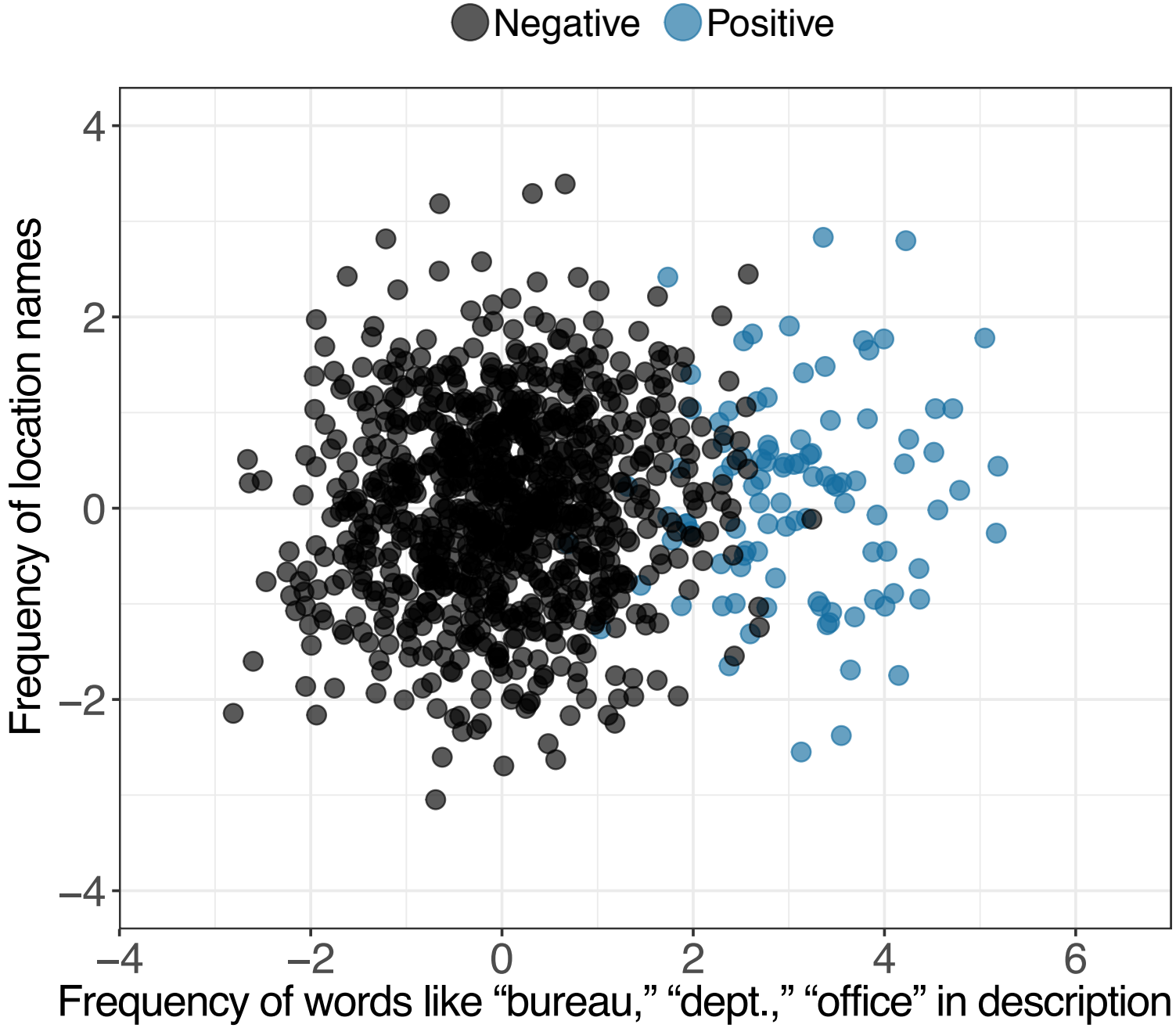


Expert/Oracle

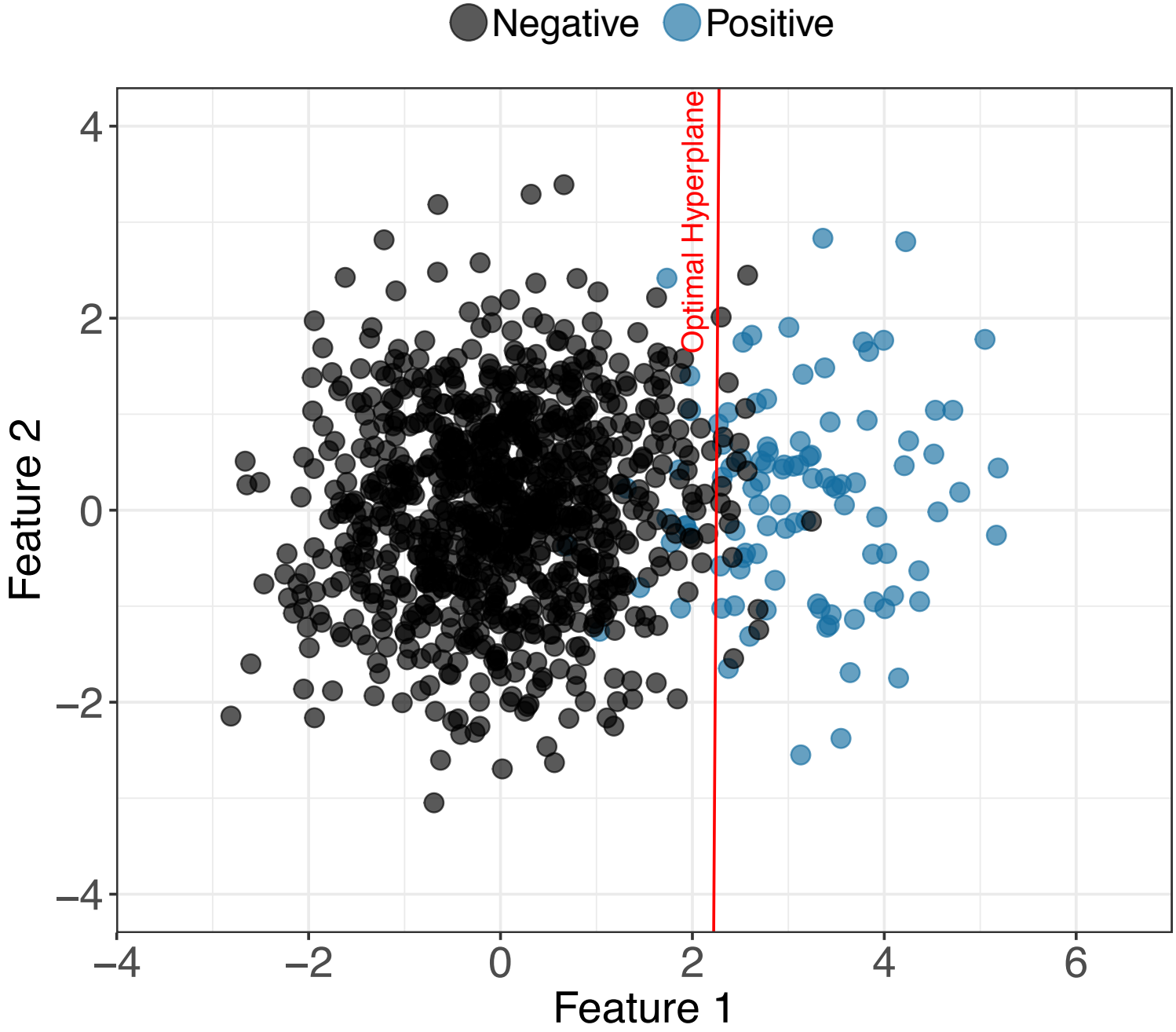
Why Active Learning?



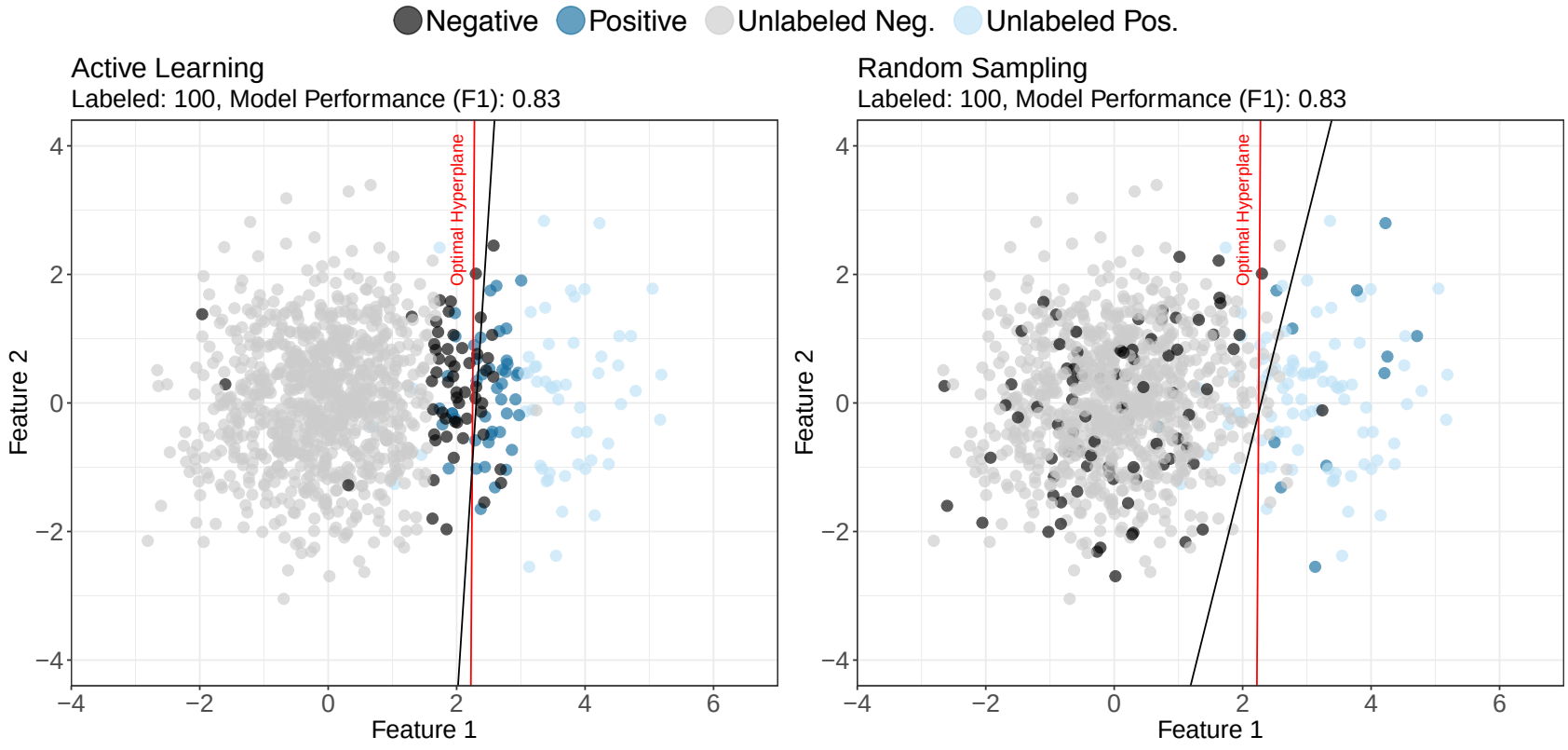
Why Active Learning?



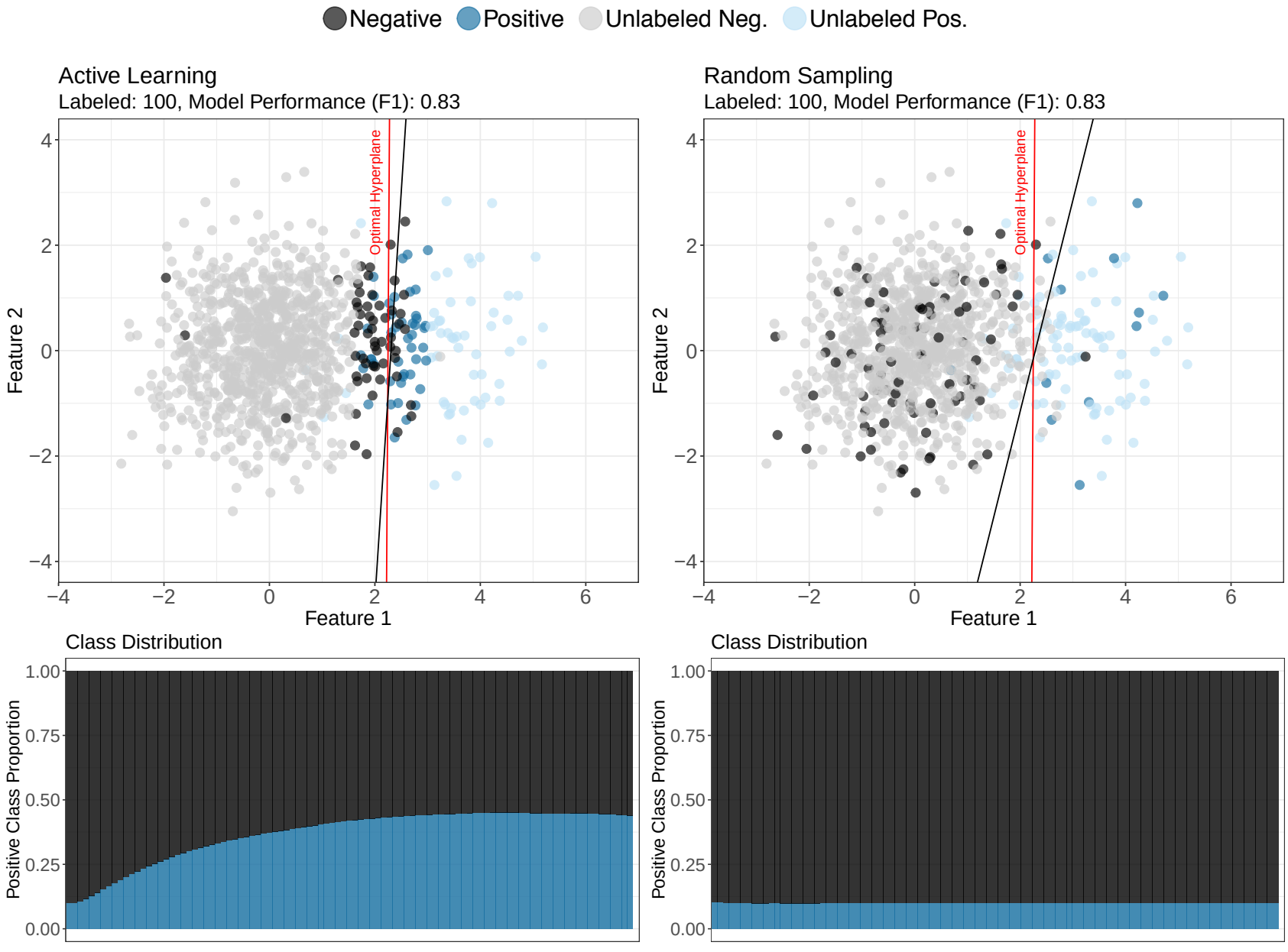
Why Active Learning?



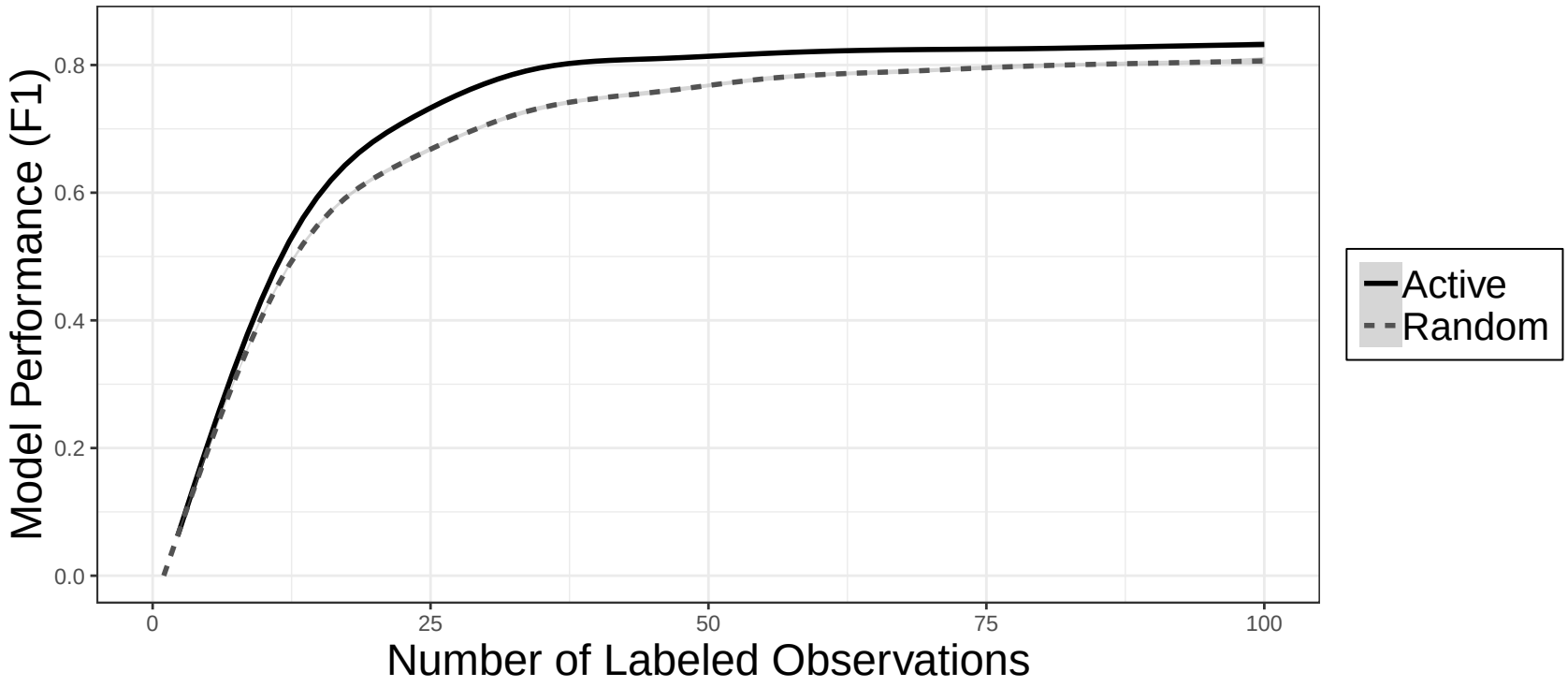
Why Active Learning?



Why Active Learning?



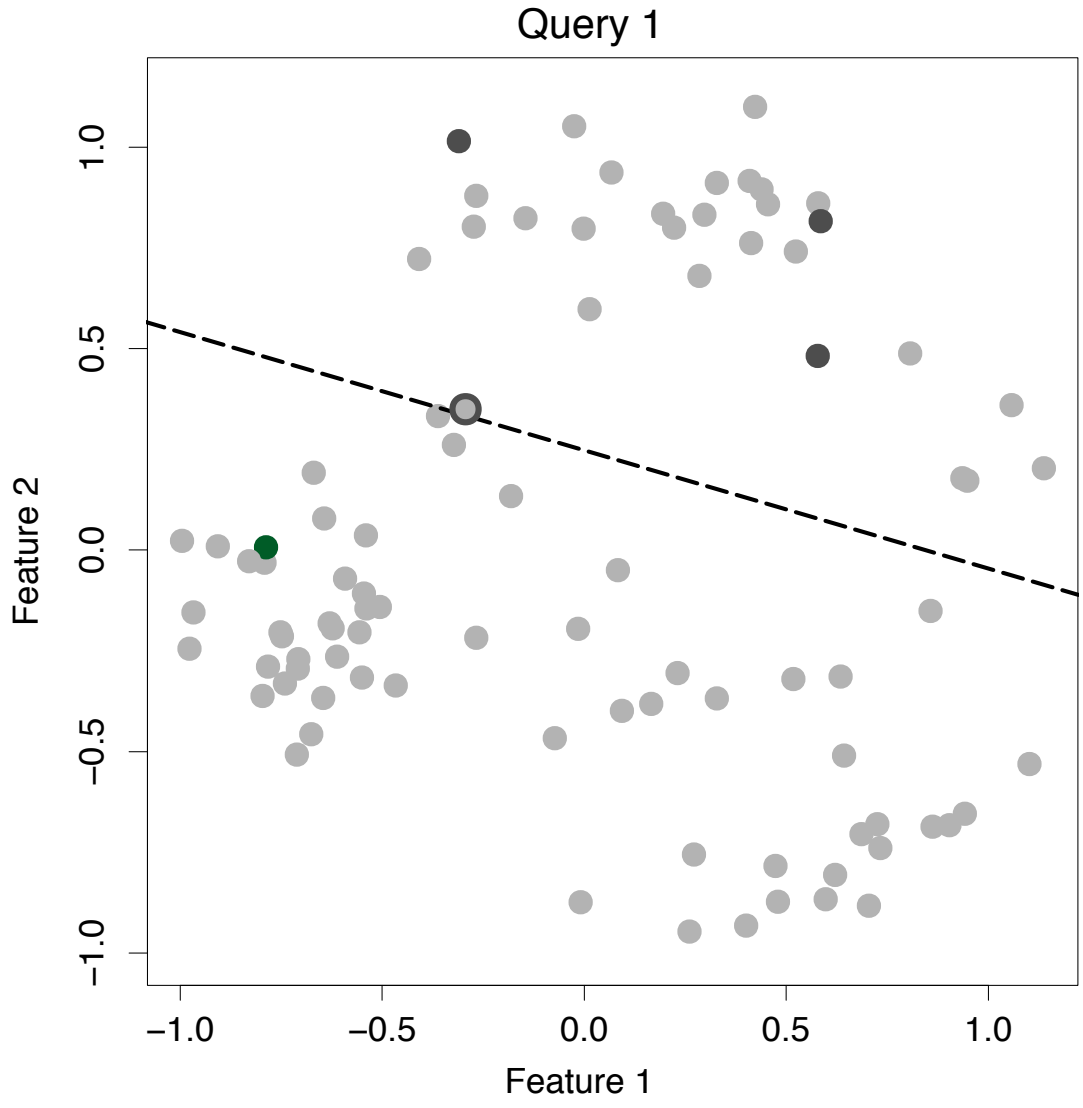
Why Active Learning?



Query Strategy

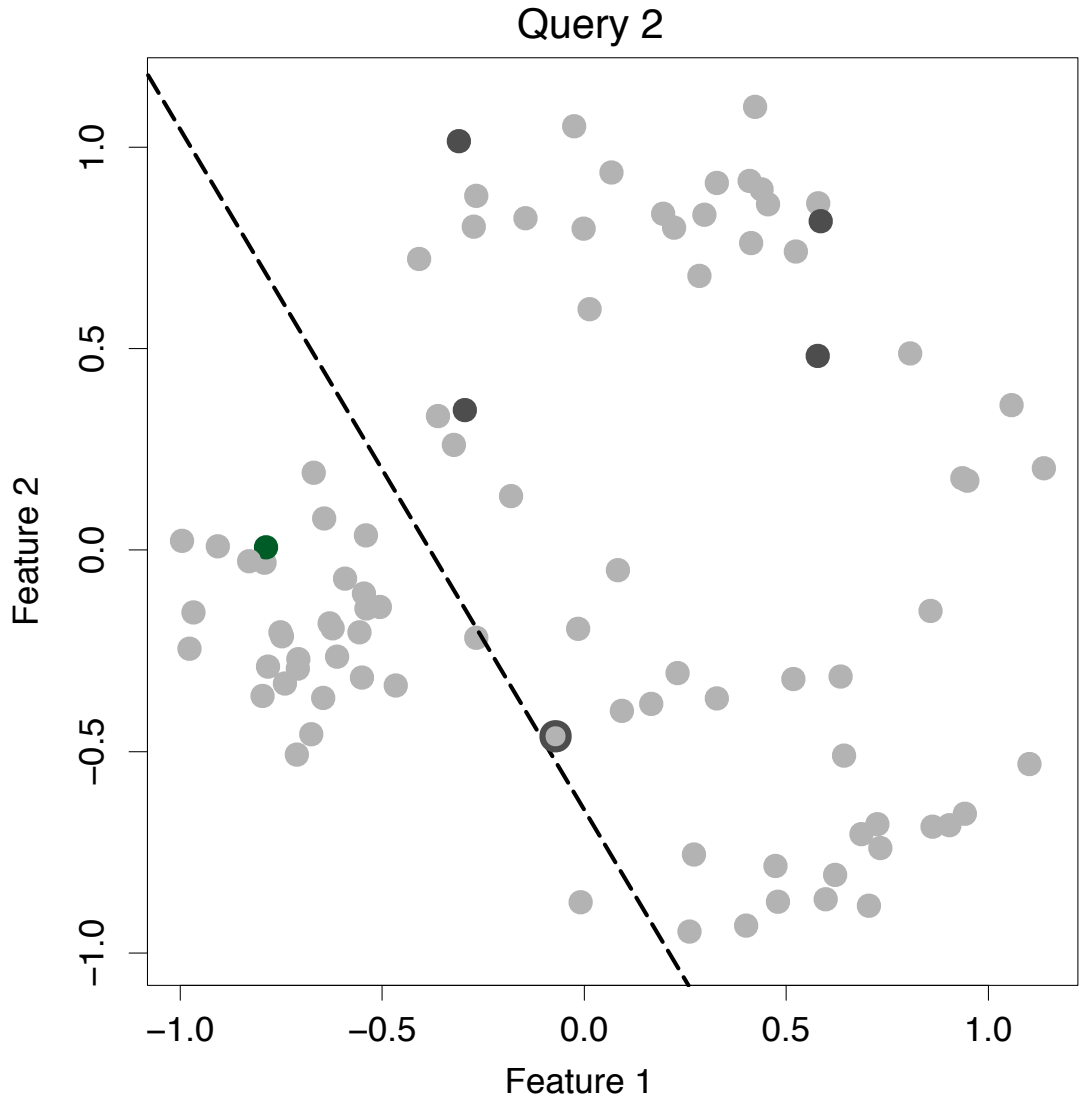
- There are many ways of calculating model uncertainty.
- Here I will discuss **margin sampling** and **query by committee**

Query Strategy: Margin Sampling



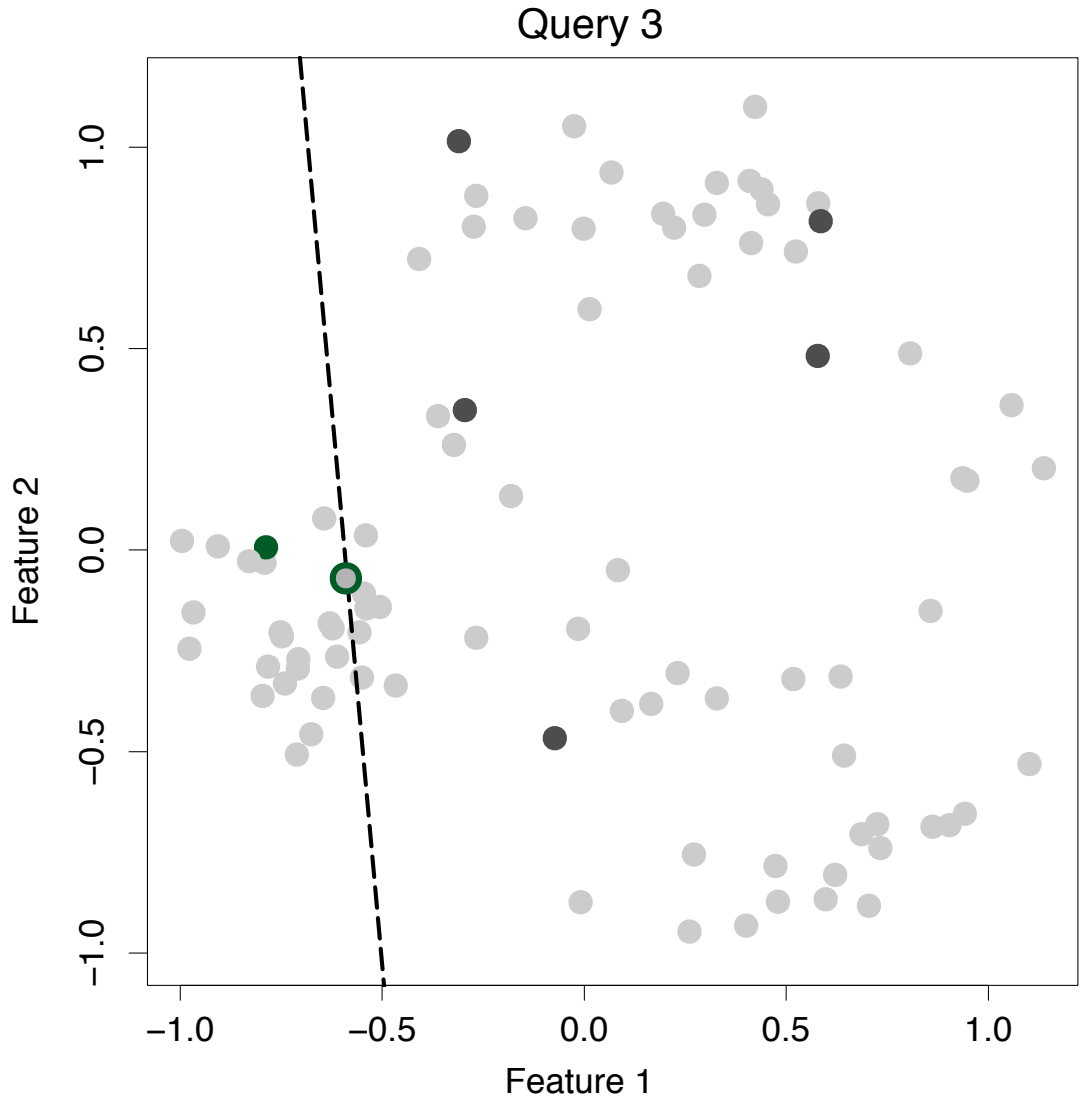
A visualization of the margin sampling algorithm.

Query Strategy: Margin Sampling



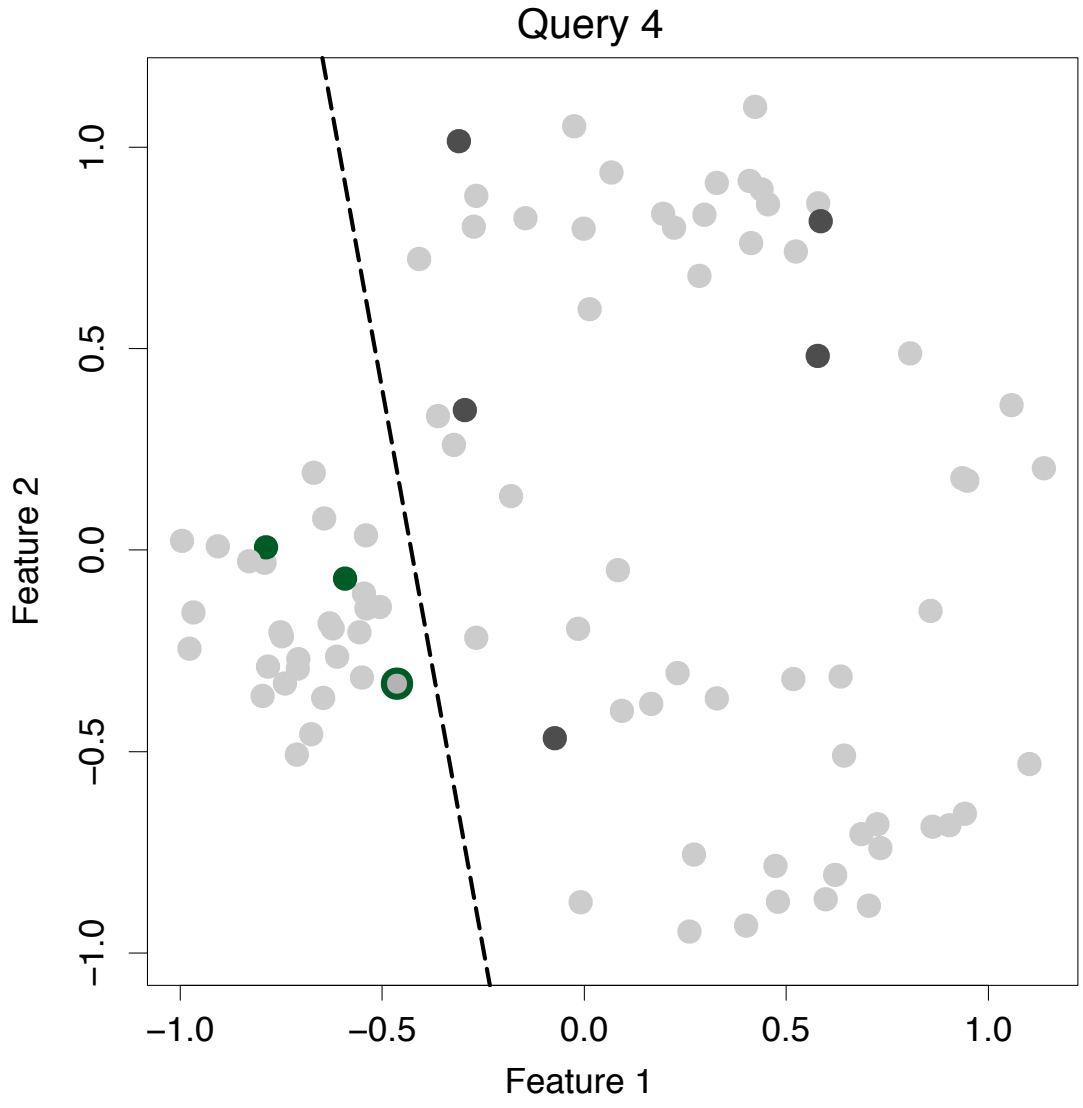
A visualization of the margin sampling algorithm.

Query Strategy: Margin Sampling



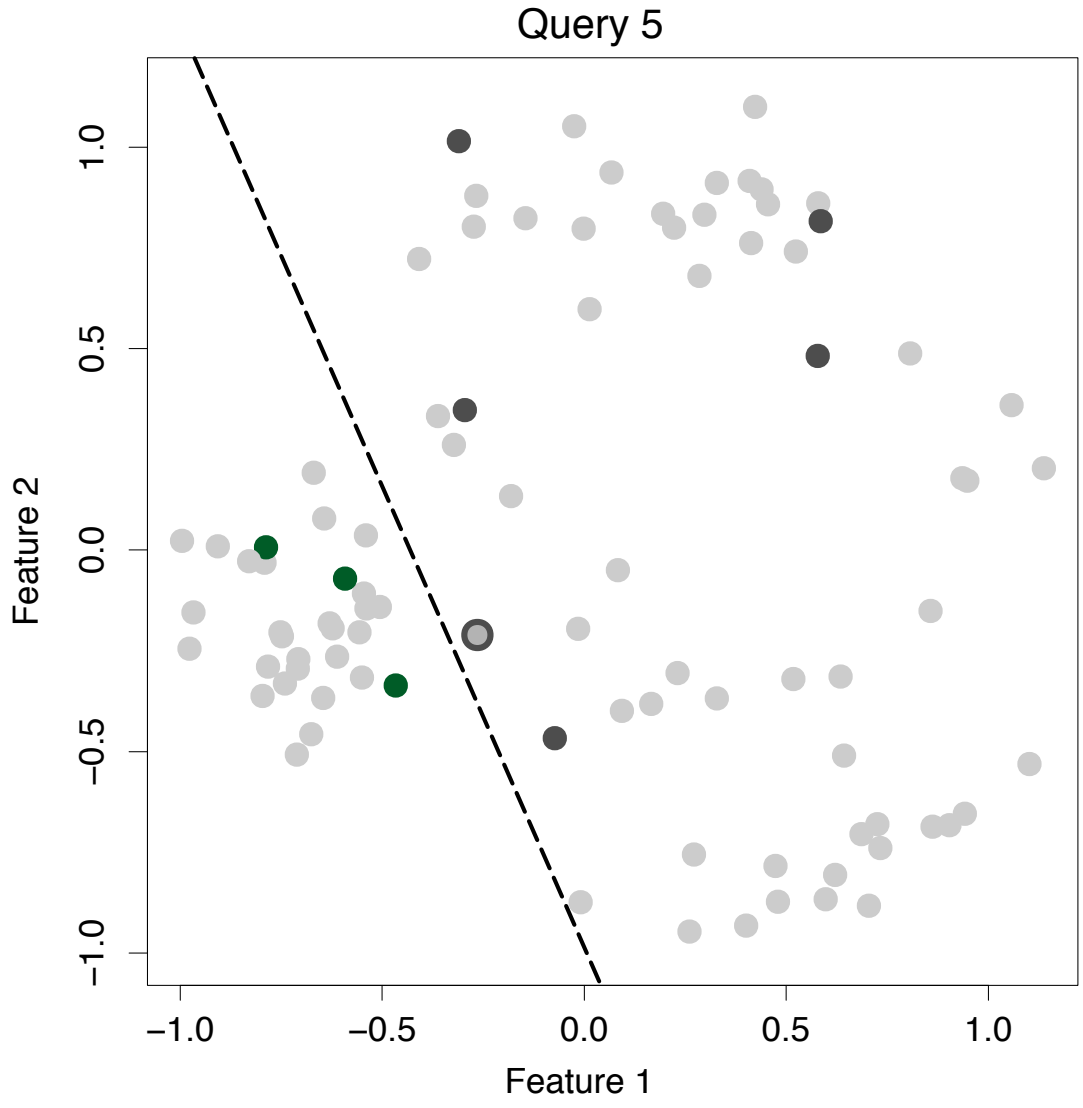
A visualization of the margin sampling algorithm.

Query Strategy: Margin Sampling



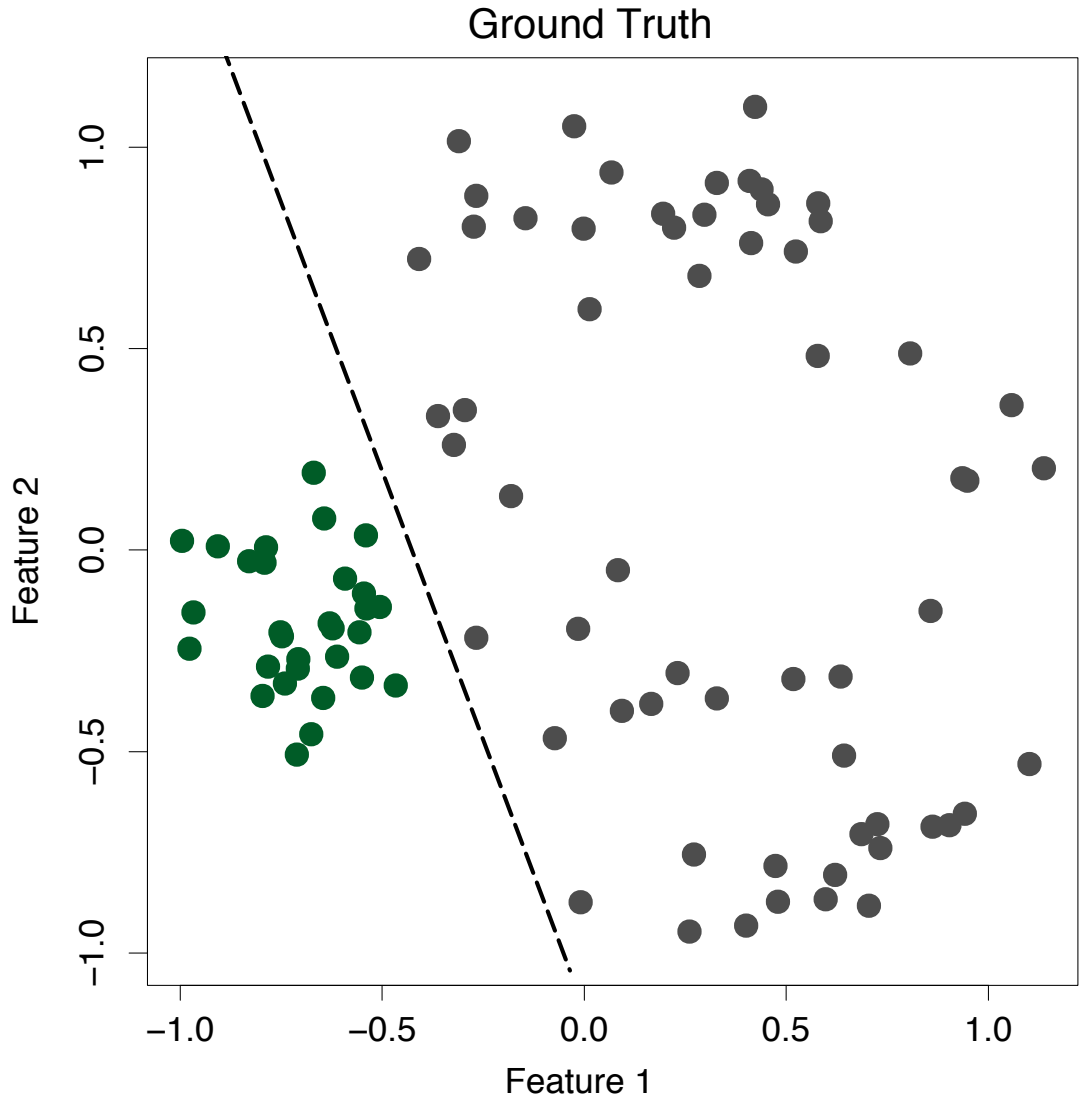
A visualization of the margin sampling algorithm.

Query Strategy: Margin Sampling



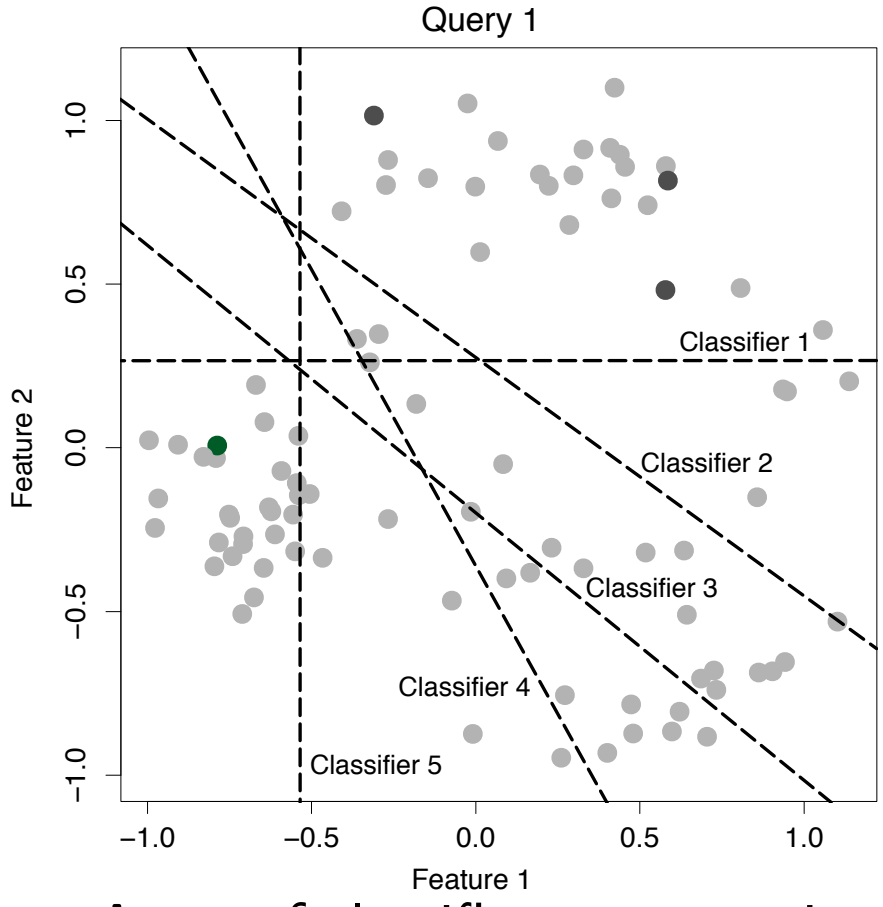
A visualization of the margin sampling algorithm.

Query Strategy: Margin Sampling



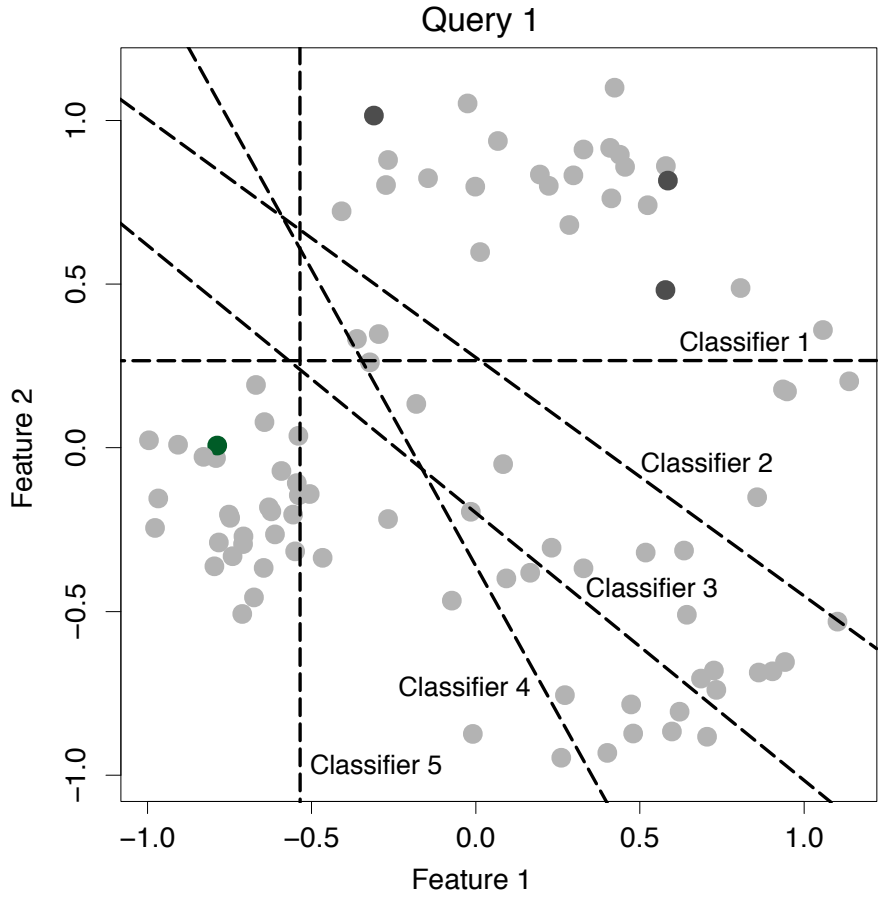
A visualization of the margin sampling algorithm.

Query Strategy: Query by Committee



- Train a **committee** of classifiers representing different hypotheses for partitioning the **version space**

Query Strategy: Query by Committee



- Train a **committee** of classifiers representing different hypotheses for partitioning the **version space**
- Select the document where committee members' predictions have the **highest disagreement**

Query Strategy: Query by Committee

Example: Ask a committee of classifiers if unlabeled documents are about refugees:

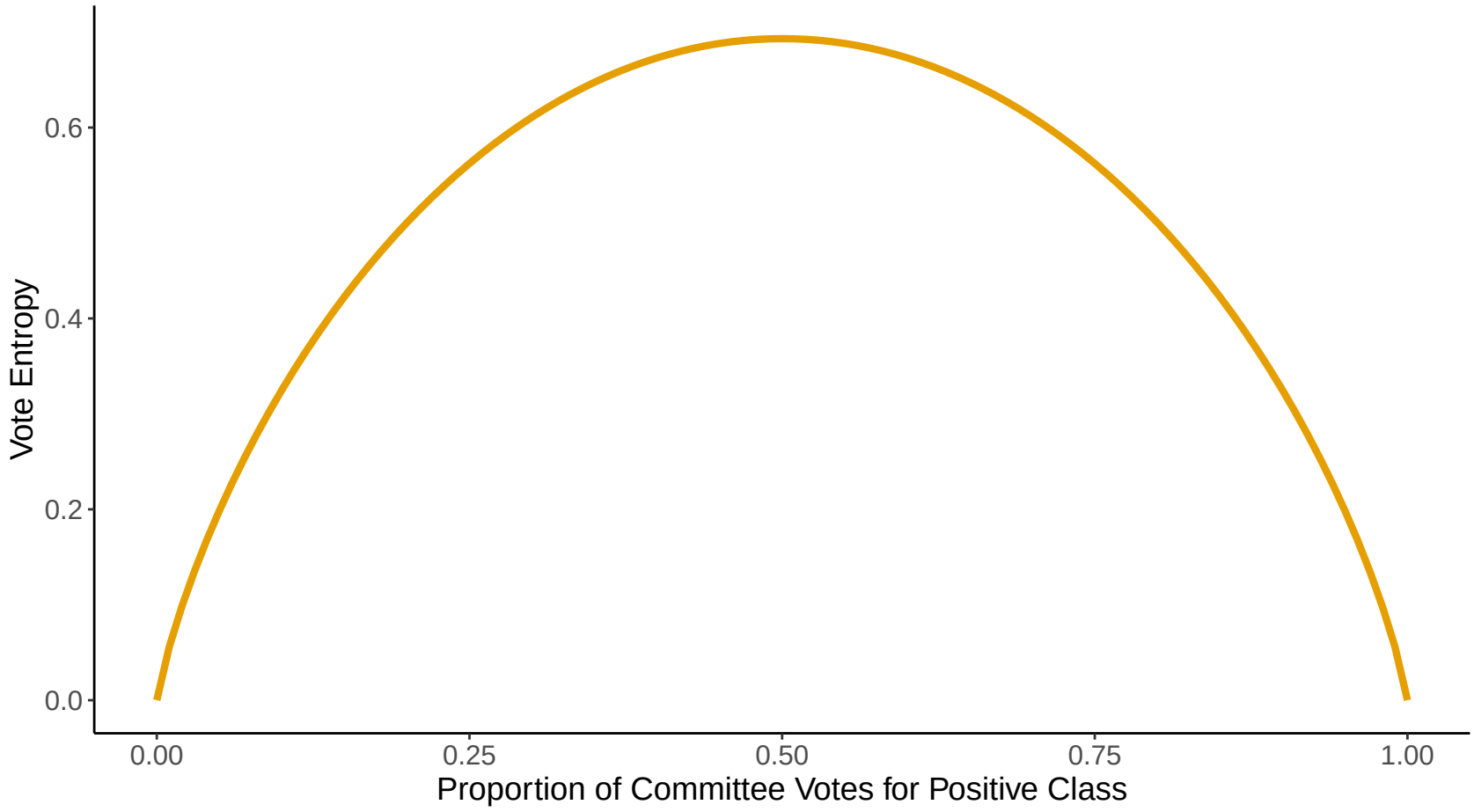
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Entropy
Document 1	YES	NO	NO	NO	NO	NO	NO	0.410
Document 2	YES	YES	YES	YES	YES	NO	NO	0.598
Document 3	YES	NO	NO	YES	YES	NO	NO	**0.683**
Document 4	NO	NO	NO	NO	NO	NO	NO	0
Document 5	YES	YES	YES	YES	YES	YES	YES	0

Calculate disagreement with **vote entropy**:

$$VE = -\sum_{i}^{k} \frac{V(y_i)}{C} log \frac{V(y_i)}{C}$$

where $V(y_i)$ is the vote count for label k, and C is committee size.

Vote entropy



Active Learning Application: Astroturfer Detection

Astroturfer Detection

Astroturfing refers to any fake or staged “grassroots” activity.

Astroturfer Detection

The Fifty Cent Party

"Netizens first coined the term 'Fifty Cent Party' to refer to undercover Internet commentators paid by the government to sway public opinion ('fifty cents' is a reference to the alleged pay received per post)."

—China Digital Times

Astroturfer Detection

Astroturfing refers to any fake or staged "grassroots" activity.

- **Research Questions:**

Astroturfer Detection

Astroturfing refers to any fake or staged "grassroots" activity.

- **Research Questions:**
  - What are Chinese government astroturfers saying?

Astroturfer Detection

Astroturfing refers to any fake or staged "grassroots" activity.

- **Research Questions:**
  - What are Chinese government astroturfers saying?
  - What are the logics of government astroturfing campaigns?

Astroturfer Detection

Astroturfing refers to any fake or staged "grassroots" activity.

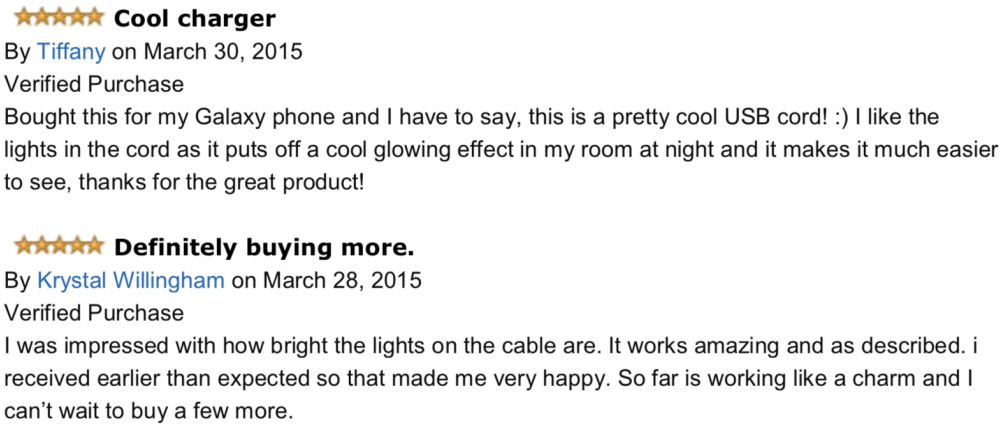
- **Research Questions:**
  - What are Chinese government astroturfers saying?
  - What are the logics of government astroturfing campaigns?
- **Method:**

Astroturfer Detection

Astroturfing refers to any fake or staged "grassroots" activity.

- **Research Questions:**
  - What are Chinese government astroturfers saying?
  - What are the logics of government astroturfing campaigns?
- **Method:**
  - Use **metadata** rather than the comment text to match expected behavioral patterns of astroturfers.

Astroturfer Detection



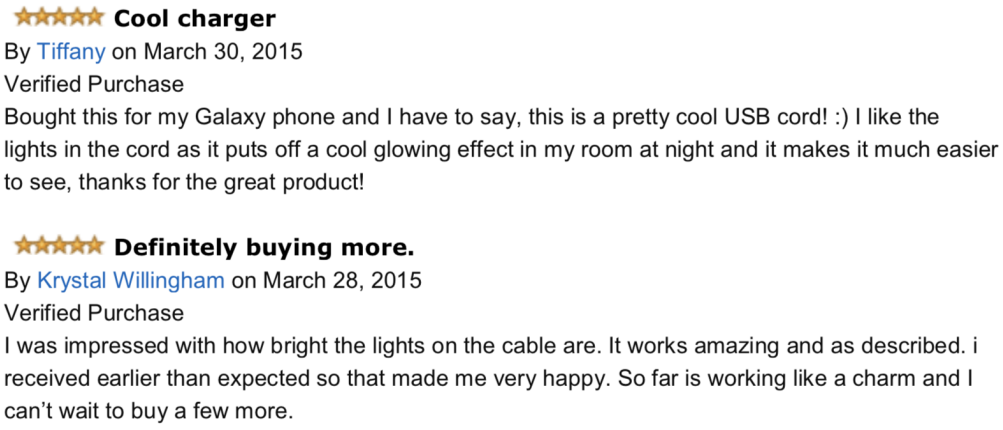
Cool charger
By Tiffany on March 30, 2015
Verified Purchase
Bought this for my Galaxy phone and I have to say, this is a pretty cool USB cord! :) I like the lights in the cord as it puts off a cool glowing effect in my room at night and it makes it much easier to see, thanks for the great product!

Definitely buying more.
By Krystal Willingham on March 28, 2015
Verified Purchase
I was impressed with how bright the lights on the cable are. It works amazing and as described. i received earlier than expected so that made me very happy. So far is working like a charm and I can't wait to buy a few more.

Amazon review astroturfing exhibit from the Amazon v. Gentile lawsuit in Washington Superior Court.

- *Research Challenge:* Need to disambiguate **government** astroturfers and **non-government** astroturfers

Astroturfer Detection



Cool charger
By Tiffany on March 30, 2015
Verified Purchase
Bought this for my Galaxy phone and I have to say, this is a pretty cool USB cord! :) I like the lights in the cord as it puts off a cool glowing effect in my room at night and it makes it much easier to see, thanks for the great product!

Definitely buying more.
By Krystal Willingham on March 28, 2015
Verified Purchase
I was impressed with how bright the lights on the cable are. It works amazing and as described. i received earlier than expected so that made me very happy. So far is working like a charm and I can't wait to buy a few more.

Amazon review astroturfing exhibit from the Amazon v. Gentile lawsuit in Washington Superior Court.

- *Research Challenge:* Need to disambiguate **government** astroturfers and **non-government** astroturfers
- *Potential Solution:* Use information from the social network of users. Do they follow **government accounts**?

Astroturfer Detection

Local government leaders are **evaluated for promotion** based on their **online influence**.

排名	微博	认证信息	传播力	服务力	互动力	认同度	总分
1	新疆地震局	新疆地震局官方微博	74.11	74.78	74.02	78.99	74.84
2	快速路交警	乌鲁木齐市城市快速路交警大队官方微博	72.75	81.73	58.03	59.14	70.56
3	平安石河子	新疆石河子市公安局官方微博	58.97	86.68	53.83	54.19	68.04
4	新疆铁路	乌鲁木齐铁路局官方微博	63.94	81.00	47.14	51.89	64.52
5	阿勒泰公安在线	新疆维吾尔自治区阿勒泰地区公安局官方微博	65.44	57.62	69.06	70.90	63.95
6	新疆反邪教	新疆维吾尔自治区防范处理邪教领导小组办公室官方微博	62.44	51.90	64.17	60.95	60.70
7	新疆平安网	新疆平安网官方微博	48.29	66.99	57.78	52.21	55.27
8	新疆消防	新疆消防总队官方微博	55.98	60.89	49.51	39.97	54.40
9	和田网警巡查执法	新疆和田地区公安局网络安全保卫支队官方微博	56.68	61.78	48.83	27.95	53.49
10	昌吉消防支队	新疆昌吉州公安消防支队官方微博	55.71	60.03	45.57	43.96	53.22

Central government rankings of **Weibo** accounts in Xinjiang Province

Astroturfer Detection

Local government leaders are **evaluated for promotion** based on their **online influence**.

Rank	Weibo Name	Weibo Description/Affiliation	Message Reach	Public Service	Inter-activity	Public Acceptance	Overall Score
1	Xinjiang Earthquake Administration	Official Weibo Account of the Xinjiang Earthquake Administration	74.11	74.78	74.02	78.99	74.84
2	Rapid Road Traffic Police	Official Weibo Account of the Urumqi City Rapid Road Traffic Police	72.75	81.73	58.03	59.14	70.56
3	Peaceful Shihezi	Official Weibo Account of the Shihezi City Public Security Bureau	58.97	86.68	53.83	54.19	68.04
4	Xinjiang Railways	Official Weibo Account of the Urumqi Railway Administration	63.94	81.00	47.14	51.89	64.52
5	Altay Online Public Security	Official Weibo Account of the Xinjiang Uighur Autonomous Region Altay Public Security Bureau	65.44	57.62	69.06	70.90	63.95
6	Xinjiang Anti-Cult	Official Weibo of the Office of the Leading Group for the Prevention and Treatment of Cults in Xinjiang Uygur Autonomous Region	62.44	51.90	64.17	60.95	60.70
7	Peaceful Xinjiang Online	Official Weibo Account of the Xinjiang Provincial Public Security Bureau	48.29	66.99	57.78	52.21	55.27
8	Xinjiang Fire Corps	Official Weibo Account of the Xinjiang Fire Corps	55.98	60.89	49.51	39.97	54.40
9	Hetian Internet Police Inspection and Law Enforcement	Official Weibo Account of the Xinjiang Hetian District Public Security Bureau Network Security Detachment	56.68	61.78	48.83	27.95	53.49
10	Changji Fire Brigade	Official Weibo Account of the Xinjiang Changji Prefecture Public Security Fire Brigade	55.71	60.03	45.57	43.96	53.22

Central government rankings of **Weibo** accounts in Xinjiang Province

Astroturfer Detection

Bureaucrats are often **required to follow the account** of the bureaucracy at which they are employed/affiliated

章贡区教育局关注“章贡发布”政务微博统计表

序号	学校（单位）	姓名	微博名称	是否已关注
1	赣七中			已关注
2	赣七中			已关注
3	赣七中			已关注
4	赣七中			已关注
5	赣七中			已关注
6	赣州市嵯峨寺小学			已关注

Document from a local propaganda department email leak.

Astroturfer Detection

Bureaucrats are often **required to follow the account** of the bureaucracy at which they are employed/affiliated

Zhanggong District Dept. of Education
Followers of Zhanggong Propaganda Department Weibo Account

Index	School (Work Unit)	Name	Weibo Username	Have they followed?
1	Ganzhou No.7 Middle School	[redacted]	[redacted]	Yes
2	Ganzhou No.7 Middle School	[redacted]	[redacted]	Yes
3	Ganzhou No.7 Middle School	[redacted]	[redacted]	Yes
4	Ganzhou No.7 Middle School	[redacted]	[redacted]	Yes
5	Ganzhou No.7 Middle School	[redacted]	[redacted]	Yes
6	Ganzhou Cuo'e Temple Elementary School	[redacted]	[redacted]	Yes

Document from a local propaganda department email leak.

Astroturfer Detection

- Need to automatically **classify** Weibo accounts as government and non-government.

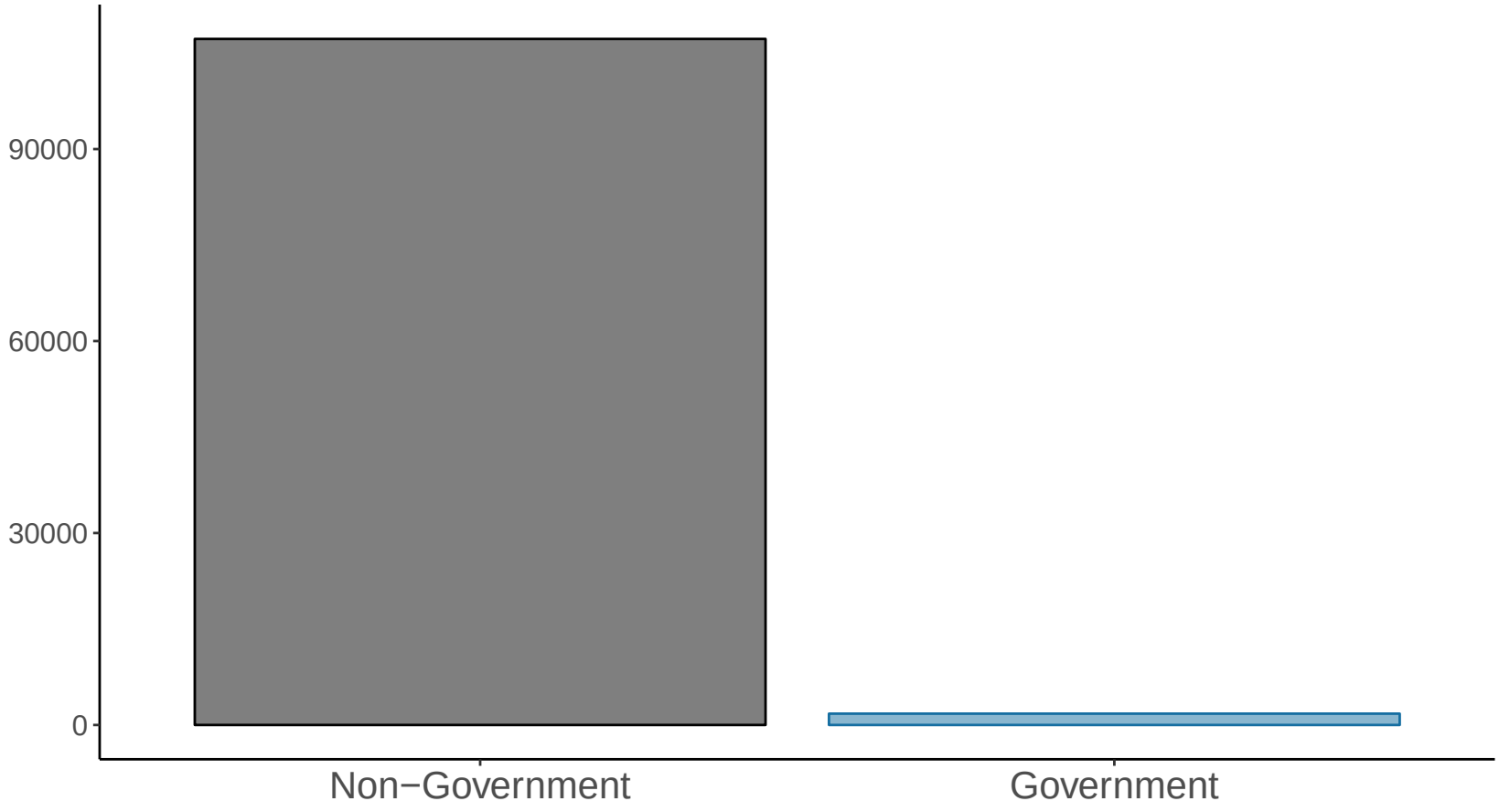
Astroturfer Detection

- Need to automatically **classify** Weibo accounts as government and non-government.
- Build a **training set** of Weibo accounts using active learning.

Astroturfer Detection

- Need to automatically **classify** Weibo accounts as government and non-government.
- Build a **training set** of Weibo accounts using active learning.
- Use this training sample to classify all followed accounts in a corpus of ~80 million news comments.

A Motivating Example



One major concern: the two classes for this problem are highly imbalanced.

Astroturfer Detection

Are Weibo accounts
government or **non-government**?

	Party	government	good	football	China	…
X_1	8	9	5	0	12	…
X_2	0	0	8	4	0	…
X_3	0	0	7	4	2	…
X_4	9	8	6	0	8	…
…	…	…	…	…	…	…

Astroturfer Detection

Model performance (F1): 0.93

Avatars of predicted government accounts.

Tianjin Explosion

Tianjin Explosion: Randomly Sampled Astroturfer Comments

- "The Tianjin incident showed the importance of firefighters and police. Salute to them!"

Tianjin Explosion: Randomly Sampled Astroturfer Comments

- "The Tianjin incident showed the importance of firefighters and police. Salute to them!"
- "We mourn for the young heroes, salute to these most beloved individuals."

Tianjin Explosion: Randomly Sampled Astroturfer Comments

- "The Tianjin incident showed the importance of firefighters and police. Salute to them!"
- "We mourn for the young heroes, salute to these most beloved individuals."
- "Don't believe in rumors or spread rumors. Rumors cease with the [intervention of] wise people. Let us all pray for [victims] and hope for their safety."

Tianjin Explosion: Randomly Sampled Astroturfer Comments

- "The Tianjin incident showed the importance of firefighters and police. Salute to them!"
- "We mourn for the young heroes, salute to these most beloved individuals."
- "Don't believe in rumors or spread rumors. Rumors cease with the [intervention of] wise people. Let us all pray for [victims] and hope for their safety."
- "Salute to these most beloved individuals. We mourn for the lost lives of the People's soldiers."

Tianjin Explosion: Comment Content

Ordinary Commentary			Astroturfer Commentary		
term	*English translation*	*weight*	*term*	*English translation*	*weight*
爆炸	explosion	0.048	致敬	to pay respects	0.168
捐款	donations	0.047	逝者	the dead	0.158
应该	should	0.038	消防官兵	firefighters	0.133
事故	accident	0.038	安息	rest in peace	0.119
天津	Tianjin	0.025	祈福	to send thoughts	0.108
天津港	Tianjin Port	0.025	天津	Tianjin	0.102
安全	safety	0.022	相信	to believe	0.095
政府	government	0.021	希望	hope	0.094
知道	know	0.020	消防	firefighters	0.088
生命	life	0.018	消防员	firefighters	0.082
希望	hope	0.018	英雄	heroes	0.072
责任	responsibility	0.018	加油	to cheer on	0.064
问题	problem	0.017	传谣	to spread rumors	0.062
发生	happen	0.015	默哀	silent tribute	0.060
砖家	“expert” (internet slang)	0.015	政府	government	0.061

Tianjin Explosion: Change in Topic Proportion