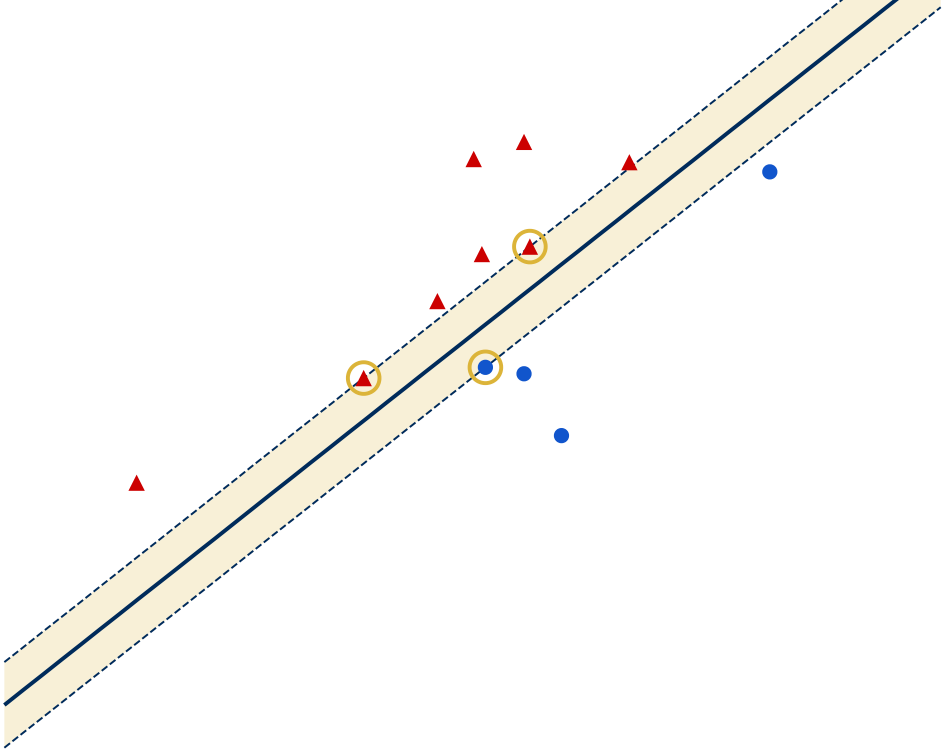


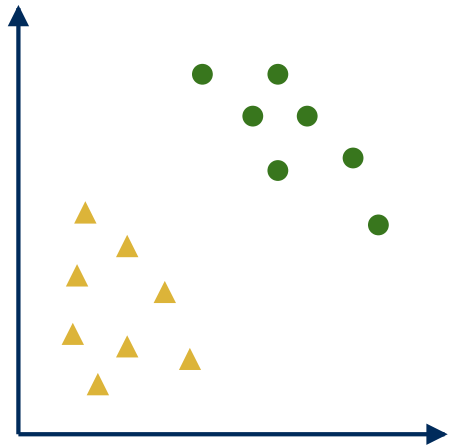
UNIT 02 · CHAPTER 9

Support Vector Machines

CSE 30124 · Introduction to Artificial Intelligence
William Theisen · University of Notre Dame

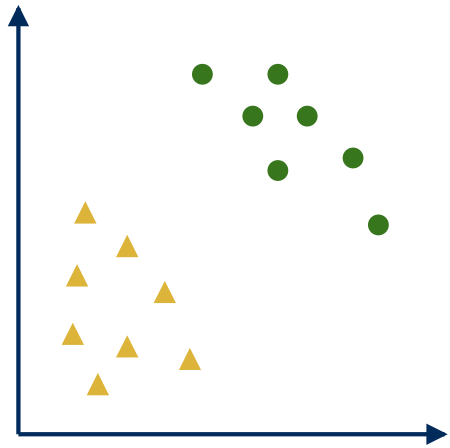
Linear Classifiers (exercise)

Given the dataset below, **where would a Perceptron draw a separating line?**

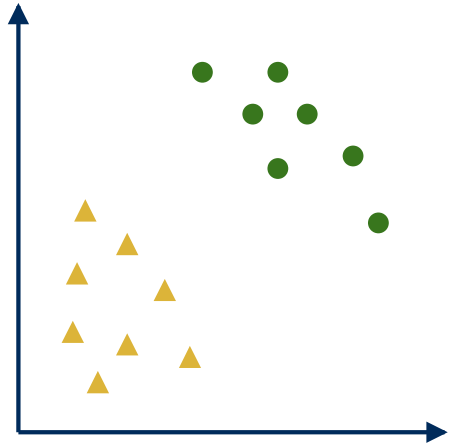


Linear Classifiers

Given the dataset below, **where would a Logistic Regression draw a separating line?**

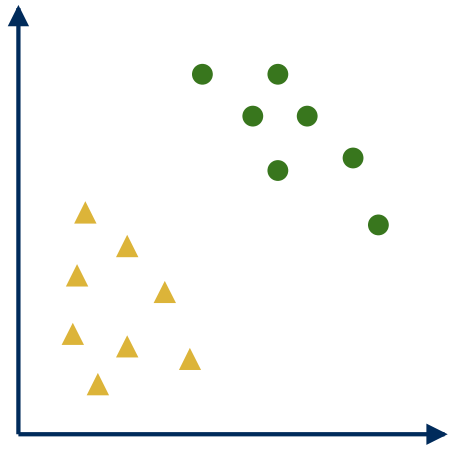


Margin Classifier



Support Vectors

Support Vectors are the points ___ or _____ the **margin** that define the **decision boundary**



Support Vector Machines (SVMs)

Support Vector Machines are a kind of **Binary Classifier** that focus on the concept of a _____ to make more robust and efficient classifiers

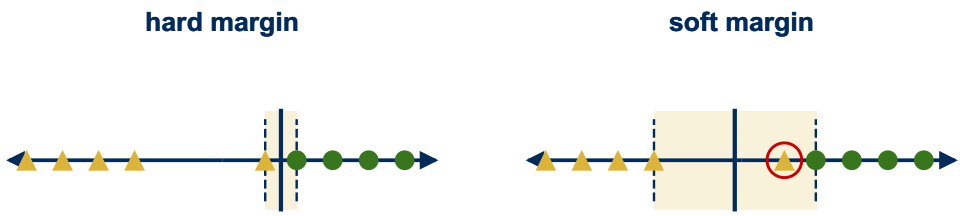
SVMs in Biology

Golub et al. (1999) measured the activity of about 7,000 genes in samples from 72 leukemia patients and asked: can we tell the two main types (AML and ALL) apart from the genes alone?

That's **thousands of features and only dozens of samples**, so infinitely many lines separate the training data perfectly and most of them are bad.

Picking the line with the **largest** _____ is a sensible way to choose, and SVMs (Furey et al., 2000) became one of the standard tools for **classifying tumors from gene activity**.

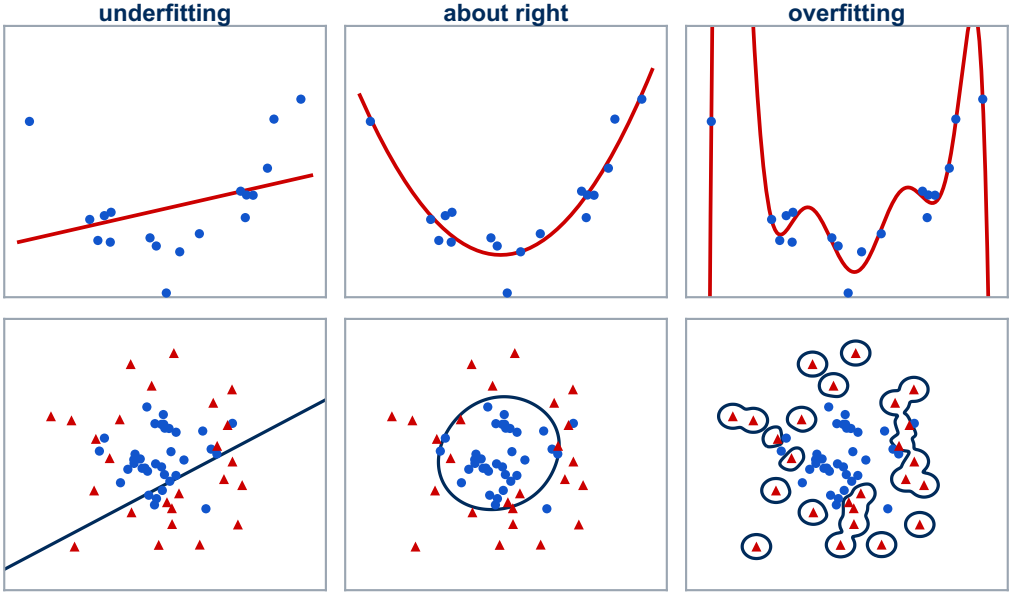
Hard-Margin vs. Soft-Margin SVMs



Hard-Margin SVMs allow no misclassifications, much like a Perceptron

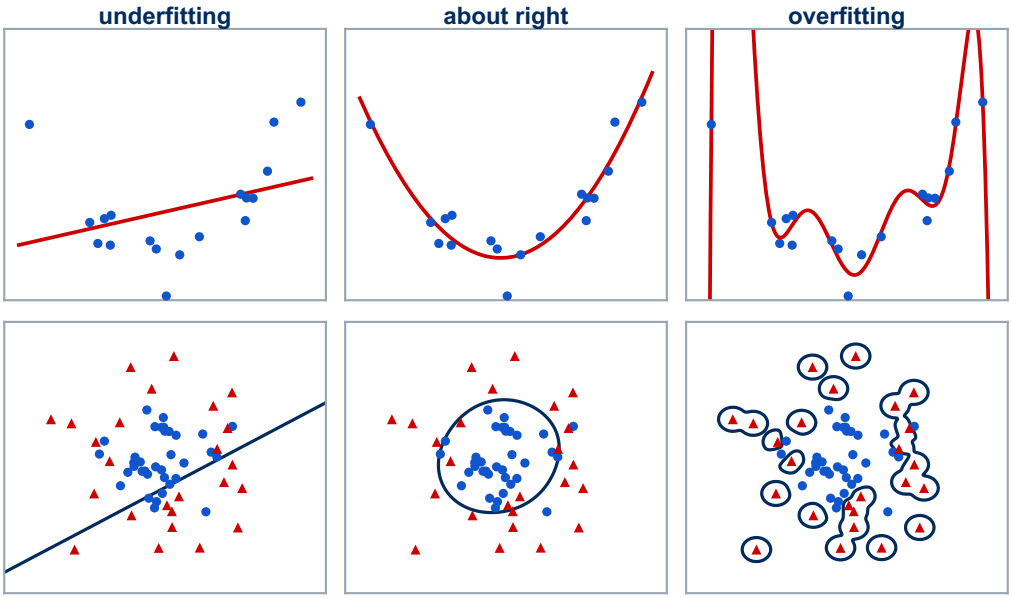
Soft-Margin SVMs allow for some misclassifications to allow for a more robust decision boundary

Underfitting and Overfitting



The bias-variance tradeoff (reading)

Underfitting and **overfitting** are _____ of having a **high bias** or a **high variance**



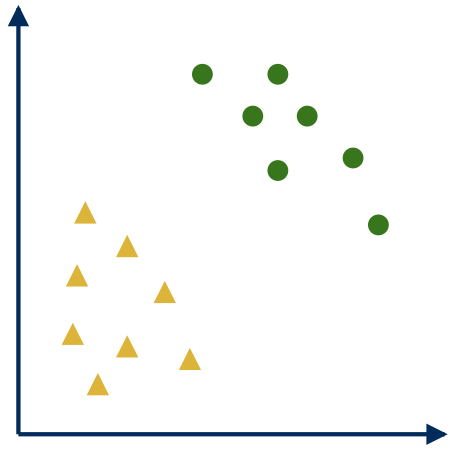
"the **bias–variance** tradeoff describes **the relationship between a model's complexity, the accuracy of its predictions during training, and the accuracy of its predictions during testing.**" (Modified from Wikipedia)

The bias-variance tradeoff (reading)



High Bias (Low Variance)	High Variance (Low Bias)

Hard-Margin Classifier Definition



We want to **minimize ||ω||** such that

all points are correctly classified and

all points are on or outside of the margin

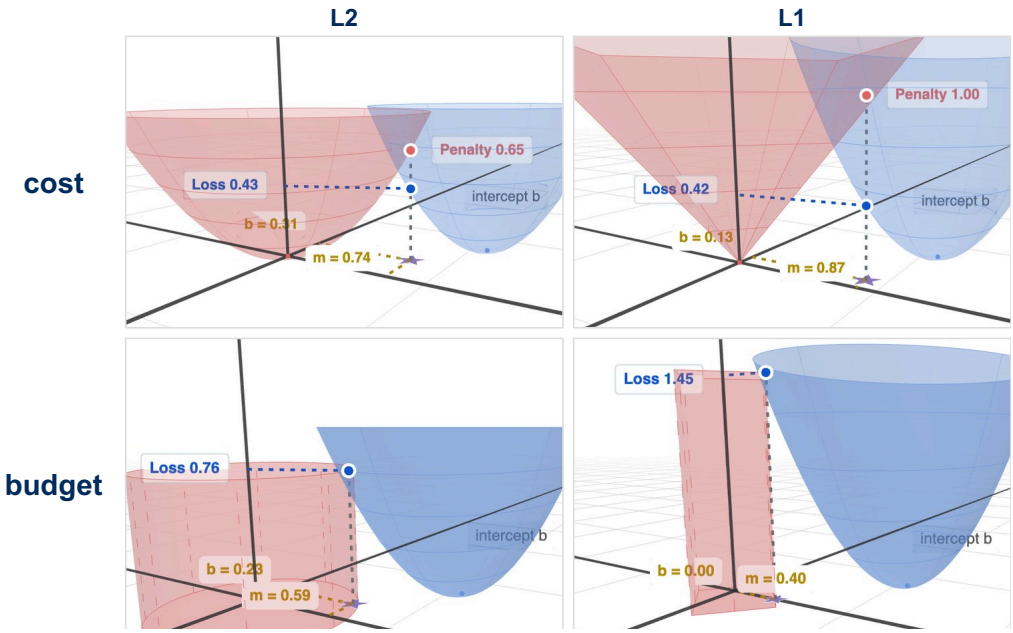
Lagrangian Multipliers

A strategy for finding the **local** maxima and **minima** of a **function subject to constraints**

Essentially, allows us to wrap the "s.t." into a **single cost/error/loss function**

Regularization

Lagrangian Multipliers are also how regularization works

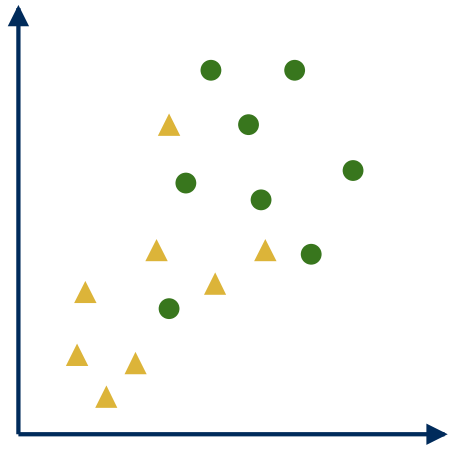


They allow us to transform the "s.t." into a **single cost/error/loss function**

Hard-Margin SVM Optimization (math)

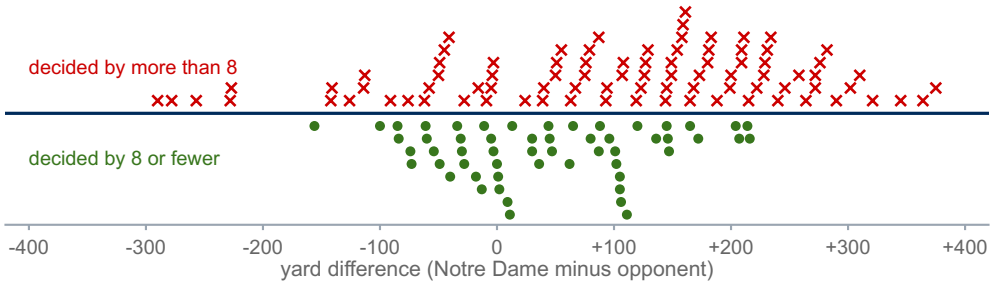
Linear Classifiers (exercise)

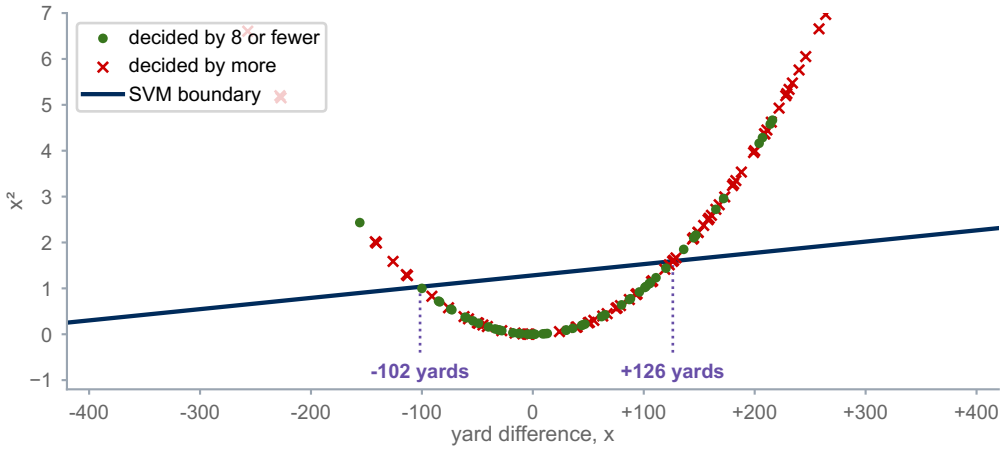
Given the dataset below, **where would a Perceptron draw a separating line?**



Soft-Margin SVMs (math)

Non-Linearly Separable Data (reading)



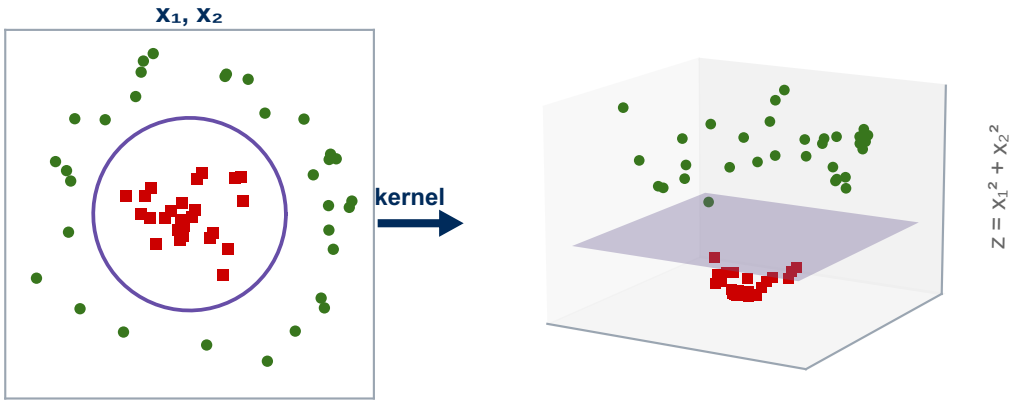


Changing your space is a very powerful idea!

The Dual-Form Equation

The Kernel Trick

What makes **SVMs** so powerful is their ability to implicitly _____ via the **"Kernel Trick"**

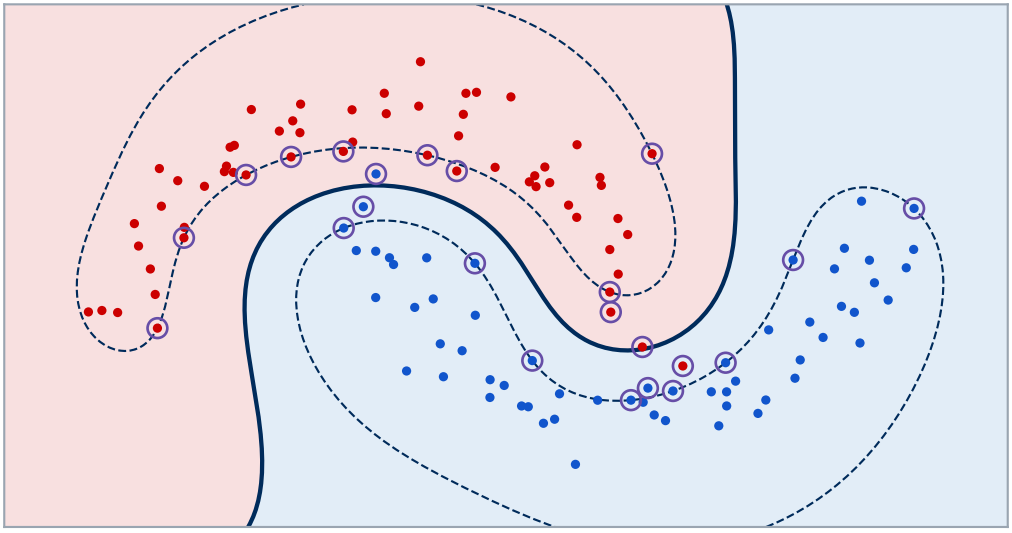


Common Kernels

There are **three common kernels**:

- _____
- _____
- _____

Visualization



Matching: Support Vector Machines

____ **Margin**	**A.** Allows some misclassifications for a more robust decision boundary
____ **Support vectors**	**B.** The gap between the decision boundary and the closest points of each class
____ **Hard-margin SVM**	**C.** Allows no misclassifications, much like a Perceptron
____ **Soft-margin SVM**	**D.** Fast and simple, but only finds a straight-line (hyperplane) boundary
___ **Lagrangian multipliers**	**E.** Builds very complex boundaries from distances; needs tuning of γ and C
____ **Kernel trick**	**F.** Computes dot products in a higher-dimensional space without transforming the data
____ **Linear kernel**	**G.** Finds curved boundaries of a chosen degree; can overfit at high degrees
____ **Polynomial kernel**	**H.** The points on or inside the margin that define the decision boundary
____ **RBF kernel**	**I.** Wrap the "s.t." constraints into a single cost/error/loss function

Practice problems

1. What makes an SVM different from a Perceptron with respect to the decision boundary?

2. What are support vectors, and why are they important in SVM?

3. When may we prefer a soft-margin classifier over a hard-margin classifier?

4. How does SVM handle non-linearly separable data? What role does the kernel function play in this?

5. What are the three most common kernels used in SVMs?

6. What is the kernel trick? (1 pt.)

7. What happens if you were to change the location of a support vector? Is it different depending on if we're using a hard-margin or soft-margin classifier? (2 pts.)