ECS 171: Machine Learning

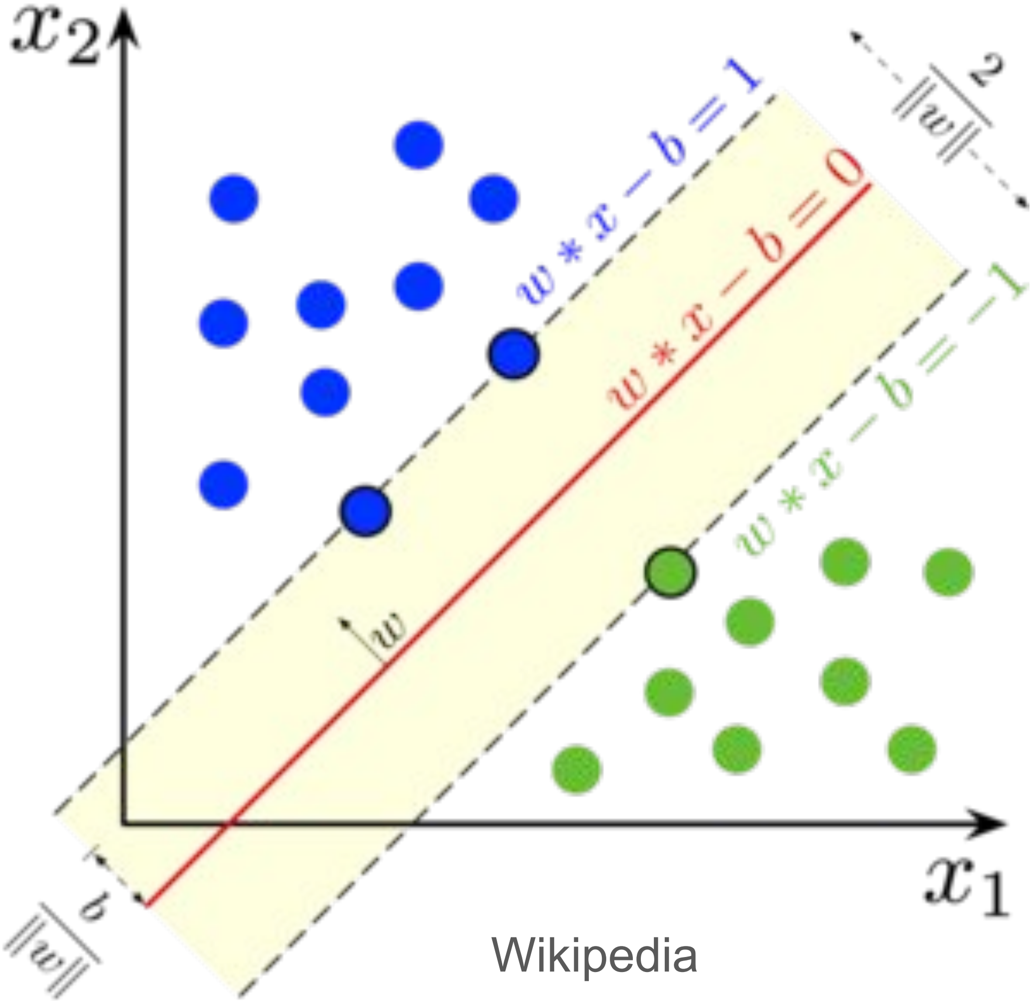
Summer 2023
Edwin Solares
easolares@ucdavis.edu
SVM Kernels

SVM Classification

We want to maximize our margin

Types:

- Hard Margins
- Soft Margins



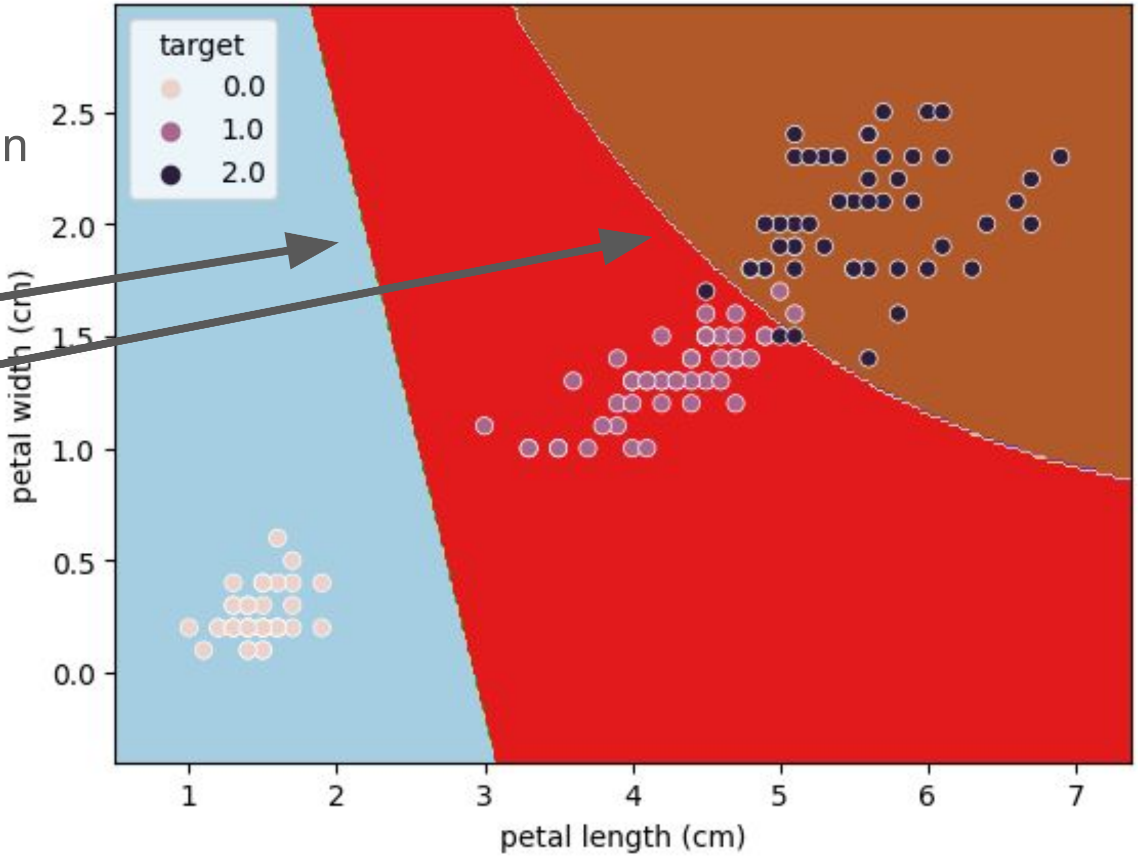
Wikipedia

SVM Classification

We want to maximize our margin

Types:

- Hard Margins
- Soft Margins



SVM Classification

Classification Boundary:

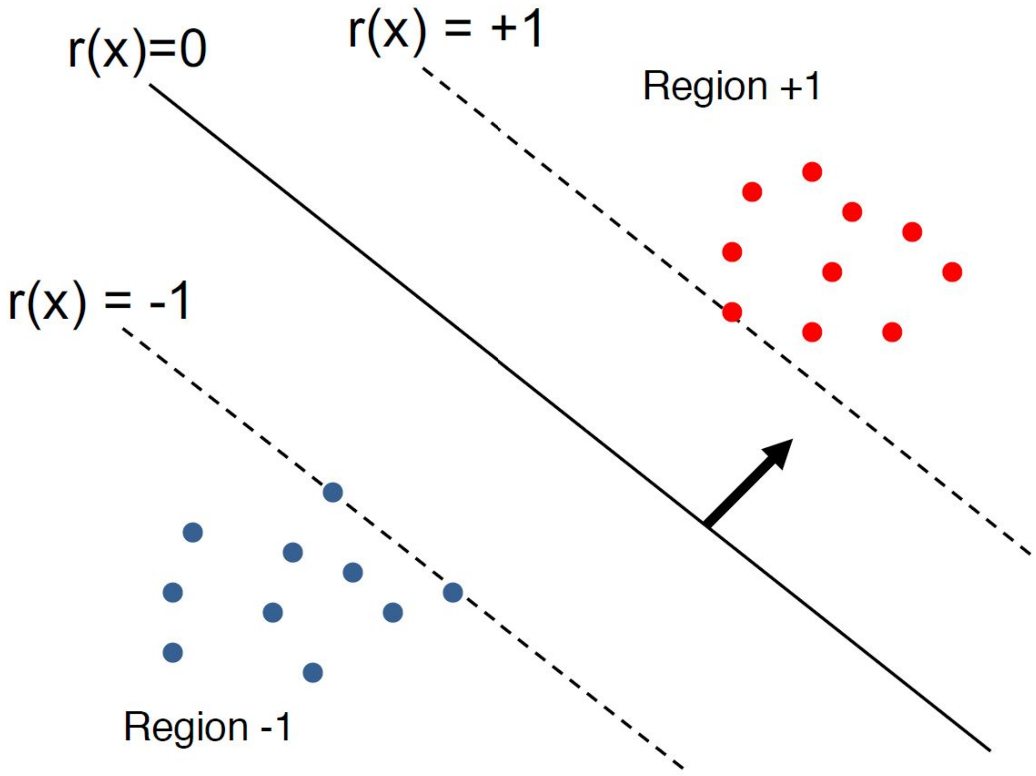
- $r(x) = wx^T + b$

Support Vectors

- Points on the Margin
- $r(x) > +1$ for the +1 Region
- $r(x) < -1$ for the -1 Region

Example of a Hard Margin

- No points within the margin



Dr. Ihler

SVM Classification

Classification Boundary:

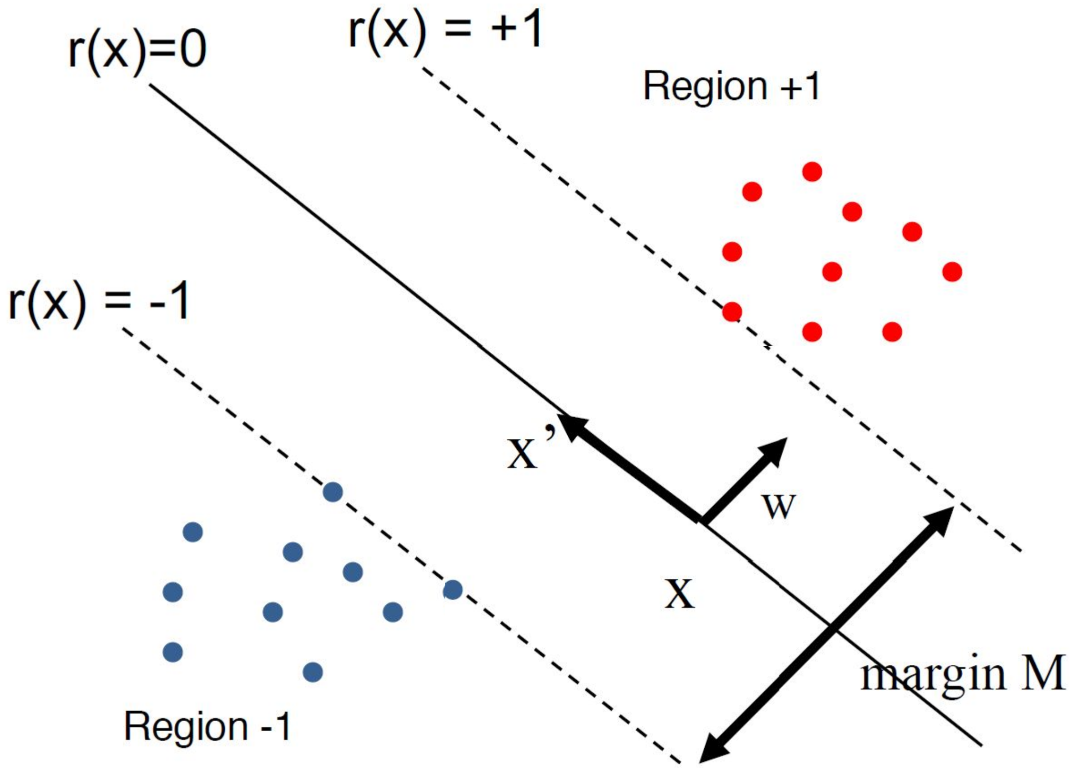
- $r(x) = wx^T + b$

Support Vectors

- Points on the Margin
- $r(x) > +1$ for the +1 Region
- $r(x) < -1$ for the -1 Region

Example of a Hard Margin

- No points within the margin



Dr. Ihler

$$w \cdot x^+ + b = +1$$

SVM Classification

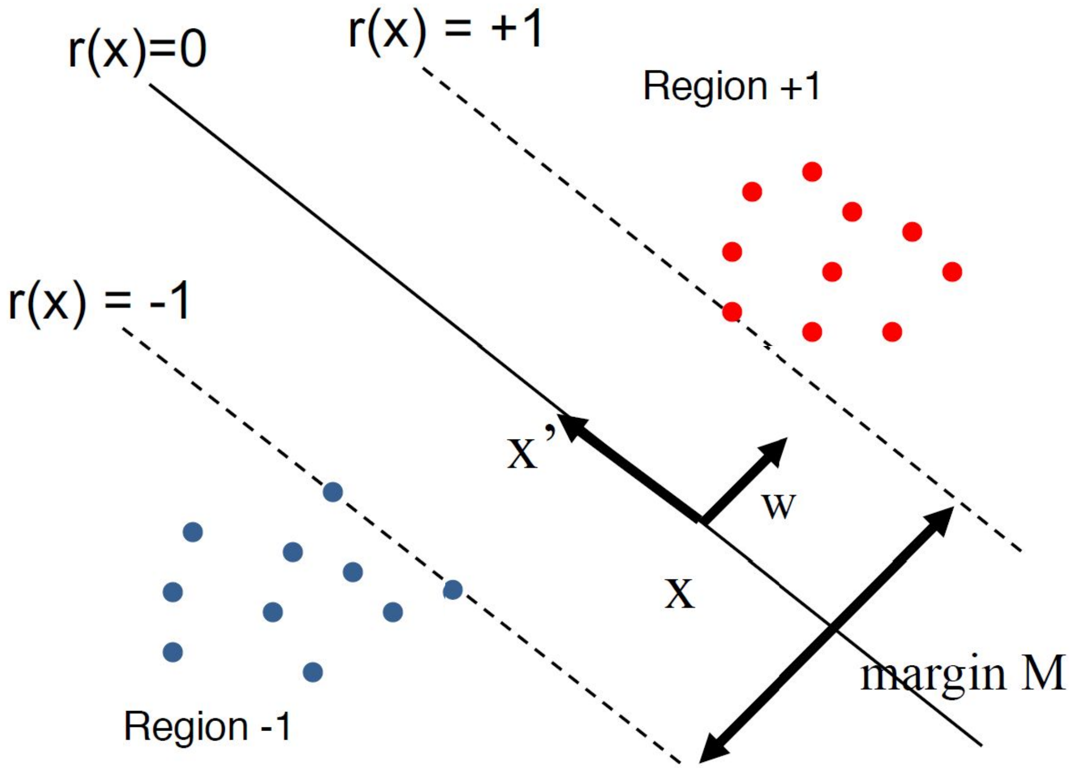
- We want to choose x+ and x- st. they are the closest to the margin

$$w \cdot (x^- + rw) + b = +1$$

$$\Rightarrow \quad r\|w\|^2 + w \cdot x^- + b = +1$$

$$\Rightarrow \quad r\|w\|^2 - 1 = +1$$

$$\Rightarrow \quad r = \frac{2}{\|w\|^2}$$



$$w \cdot x^- + b = -1$$

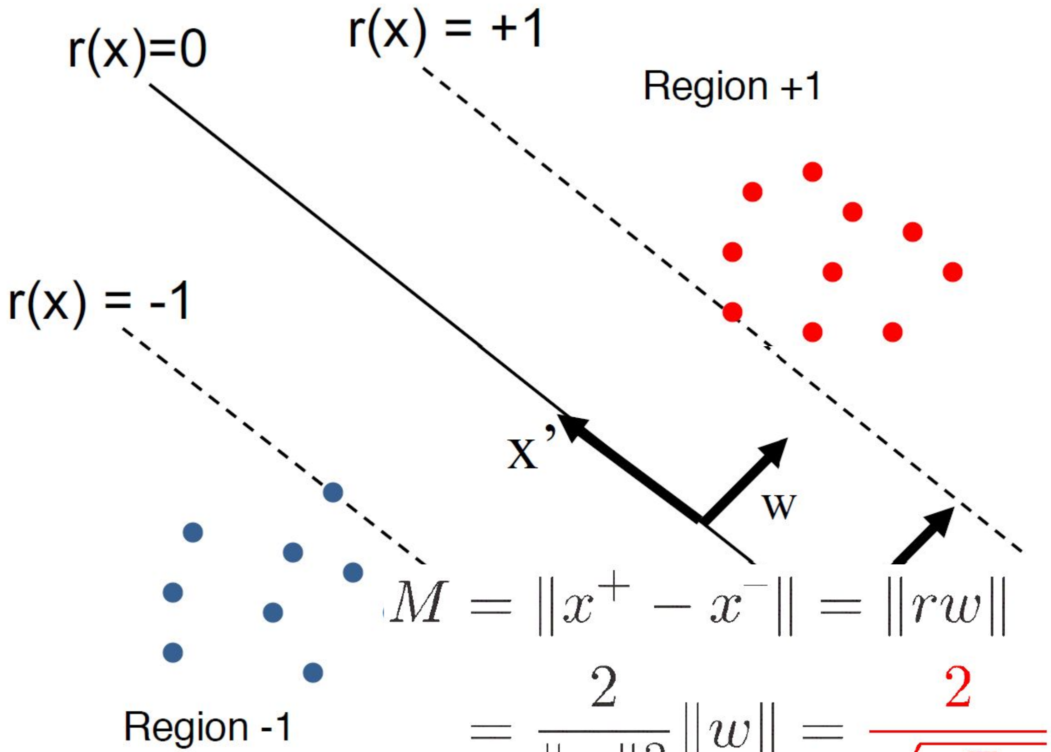
Dr. Ihler

$$w \cdot x^+ + b = +1$$

SVM Classification

- We want to choose x+ and x- st. they are the closest to the margin

$$w \cdot (x^- + rw) + b = +1$$
$$\Rightarrow \quad r\|w\|^2 + w \cdot x^- + b = +1$$
$$\Rightarrow \quad r\|w\|^2 - 1 = +1$$
$$\Rightarrow \quad r = \frac{2}{\|w\|^2}$$



$$M = \|x^+ - x^-\| = \|rw\|$$
$$= \frac{2}{\|w\|^2}\|w\| = \frac{2}{\sqrt{w^T w}}$$

$$w \cdot x^- + b = -1$$

Dr. Ihler

SVM Classification

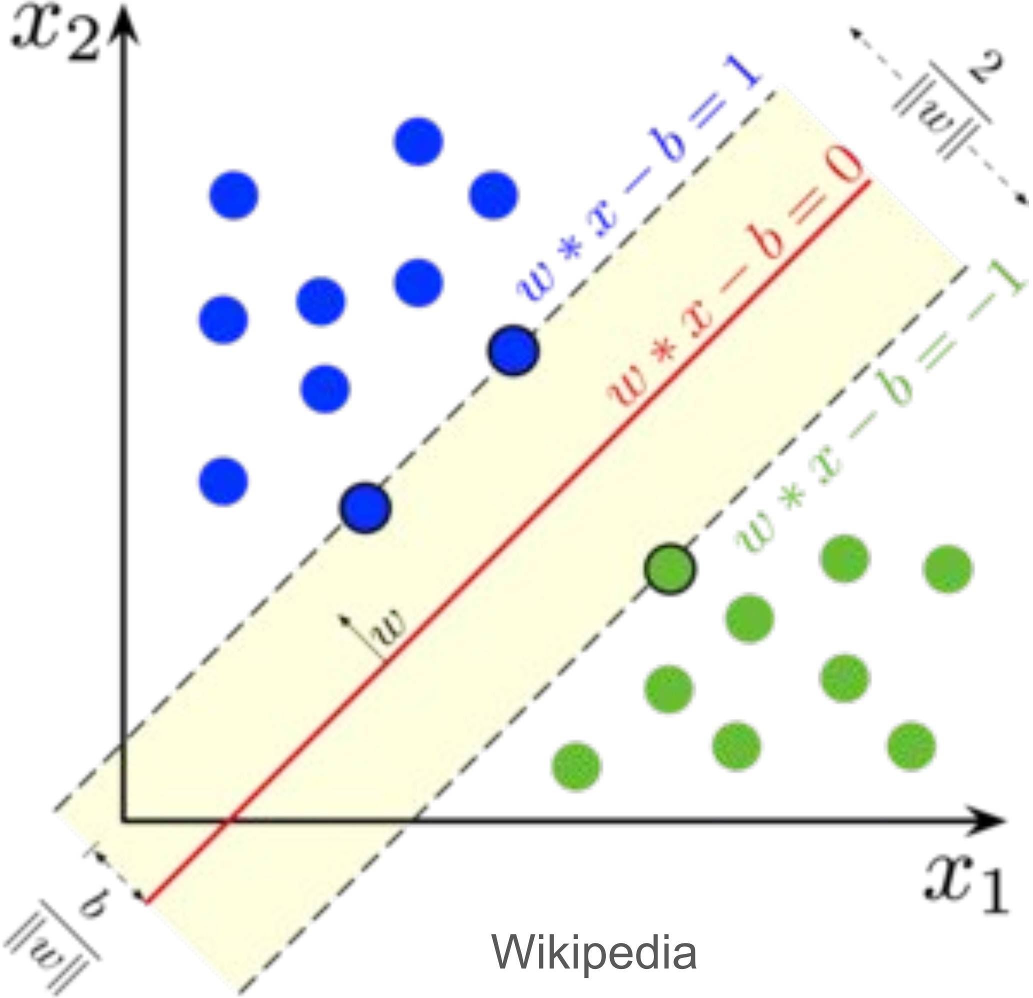
Constraints:

$y_i = +1 \rightarrow wx_i + b >= +1$

$y_i = -1 \rightarrow wx_i + b <= -1$

$$w^* = \arg\max_w \frac{2}{\sqrt{w^T w}}$$

$$w^* = \arg\min_w \sum_j w_j^2$$



Wikipedia

SVM Classification

Constraints:

$y_i = +1 \rightarrow wx_i + b >= +1$

$y_i = -1 \rightarrow wx_i + b <= -1$

$$w^* = \arg\max_w \frac{2}{\sqrt{w^T w}}$$

$$w^* = \arg\min_w \sum_j w_j^2$$

Last week's discussion:

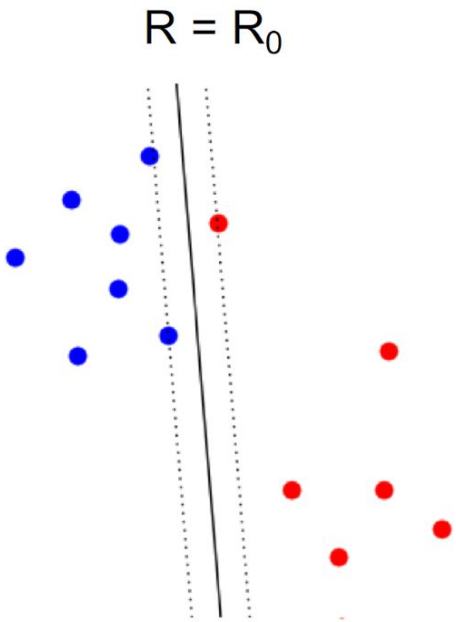
- Lagrangian optimization

Soft Margin

- Set +1 - ϵ_i
- st. ϵ >= 0 & R ∝ distance from M

$$w^* = \arg\min_{w,\epsilon} \sum_j w_j^2 + R \sum_i \epsilon^{(i)}$$

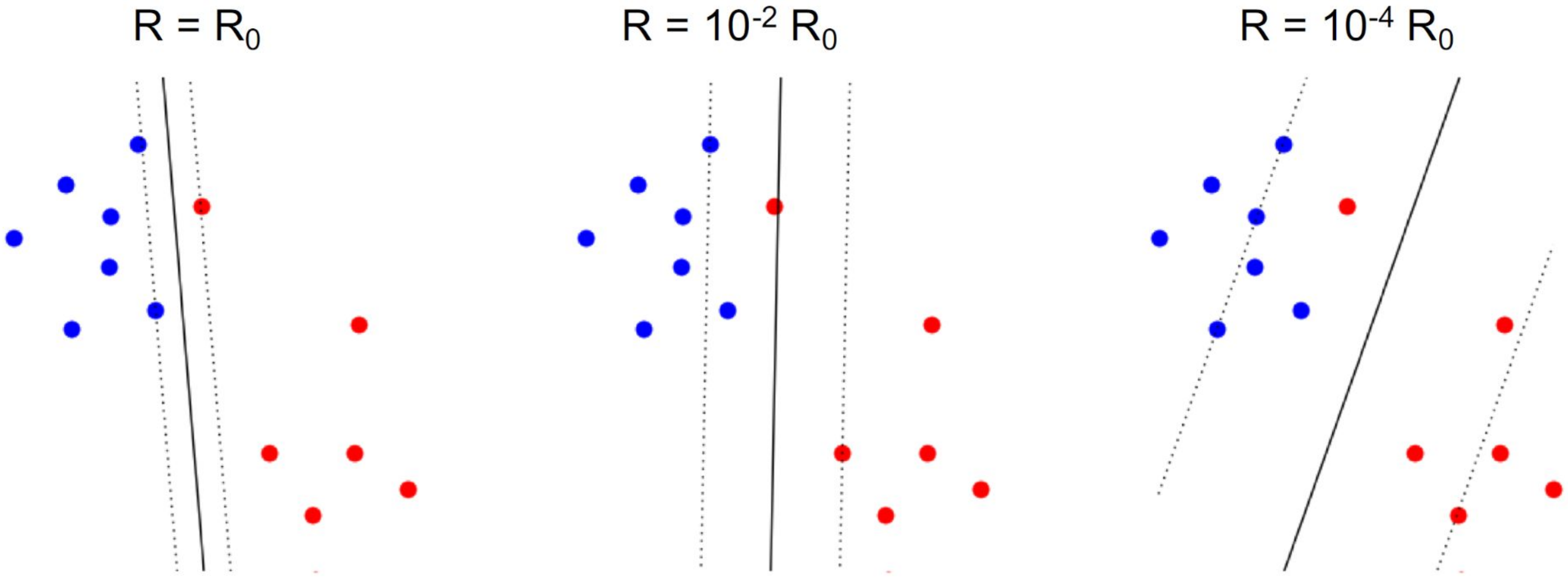
Slack Variables and Margins with Error



Hard Margin

Soft Margin

Slack Variables and Margins with Error



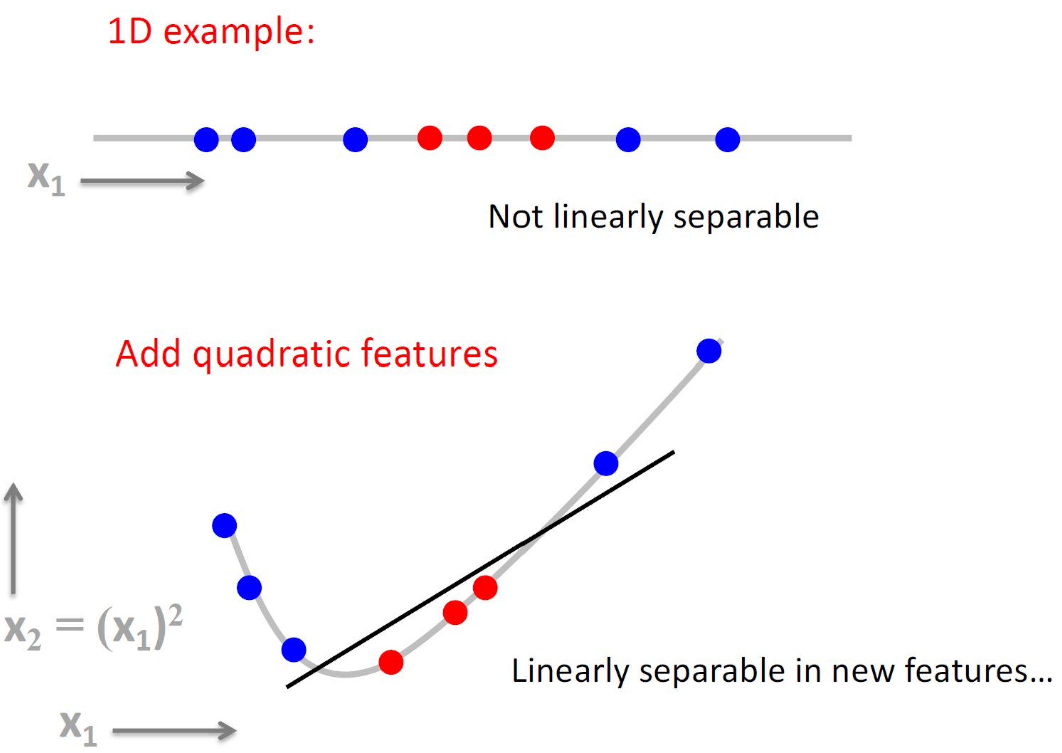
Hard Margin

Soft Margin

Very Soft Margin

Dr. Ihler

Kernels: Adding features for better classification



Kernel

Dr. Ihler

Kernels: Transforming X

$$\hat{y}(x) = \text{sign}\left[\, w \cdot \Phi(x) + b \,\right]$$

Where $\mathbf{\Phi}$(x) allows us to transform x
For example we can make $\mathbf{\Phi}$(x) = ?

- $\mathbf{\Phi}(x) = x^2$
- $\mathbf{\Phi}(x) = x^3$
- $\mathbf{\Phi}$(x) = sqroot(x)

Generalized we can define our transform as:

$$K(a, b) = (1 + \sum_j a_j b_j)^d$$

Dr. Ihler

Kernels: Transforming X

$$K(a, b) = (1 + \sum_j a_j b_j)^d$$

d is our degree and a is our *x* and *b* is our *x'* and *j* the number of expanded features
Where *r = 1* and represents our x'' and our coefficient

Simply $(a, a^2, ½) \cdot (b, b^2, ½)$ for x and x' and x''
(we have x-axis, y-axis an z-axis coordinates)
Note: when z-axis coordinates are equal we can ignore

Kernels: Transforming X

$$K(a,b) = (r + \sum_j a_j b_j)^d$$

For d = 2 and r = ½ on *K(x, x')* we get:
For d = 2 and r = ½ we get: ab + a^2b^2 *+ 1*
(a, a^2, ½) · (b, b^2, ½)

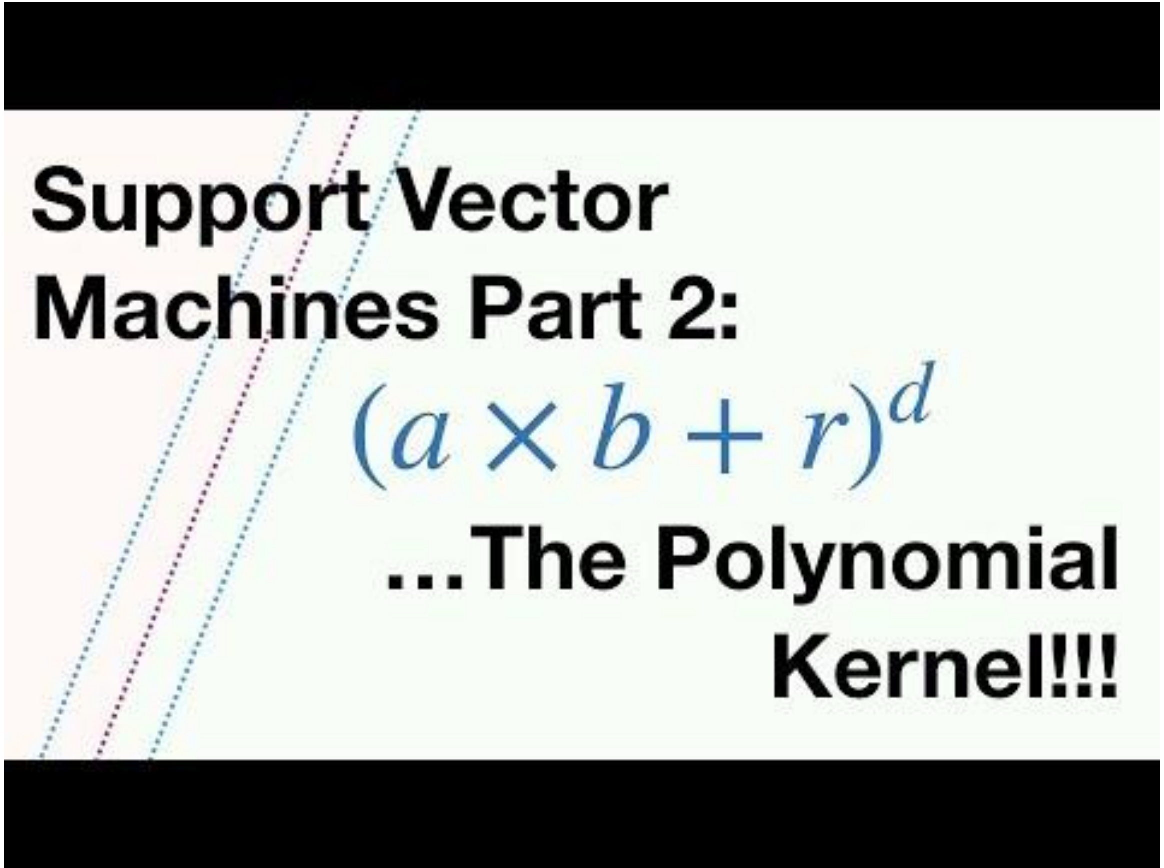
For d = 2 and r = 2 on *K(x, x')* we get:
For d = 2 and r = 2 we get: 4ab + a^2b^2 *+ 4*
(2a, a^2, 2) · (2b, b^2, 2)
Here the 2 in 2a moves our points on the x-axis

Kernels Trick

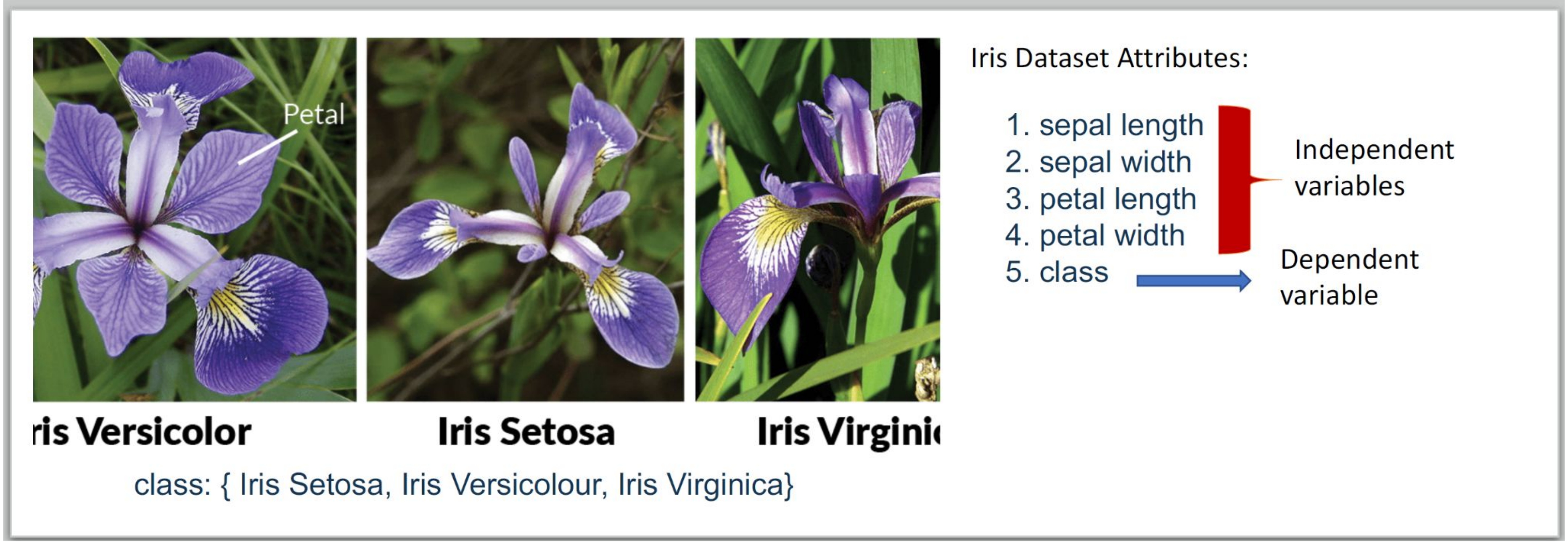
$$K(a,b) = (r + \sum_j a_j b_j)^d$$

Perform comparative analysis between different points to evaluate higher dimensional relationship. I.e. For feature vector x_1 observations $x_{1,1}$ and $x_{2,1}$, set $x_{1,1}$ to a and $x_{2,1}$ to b

Kernels Trick



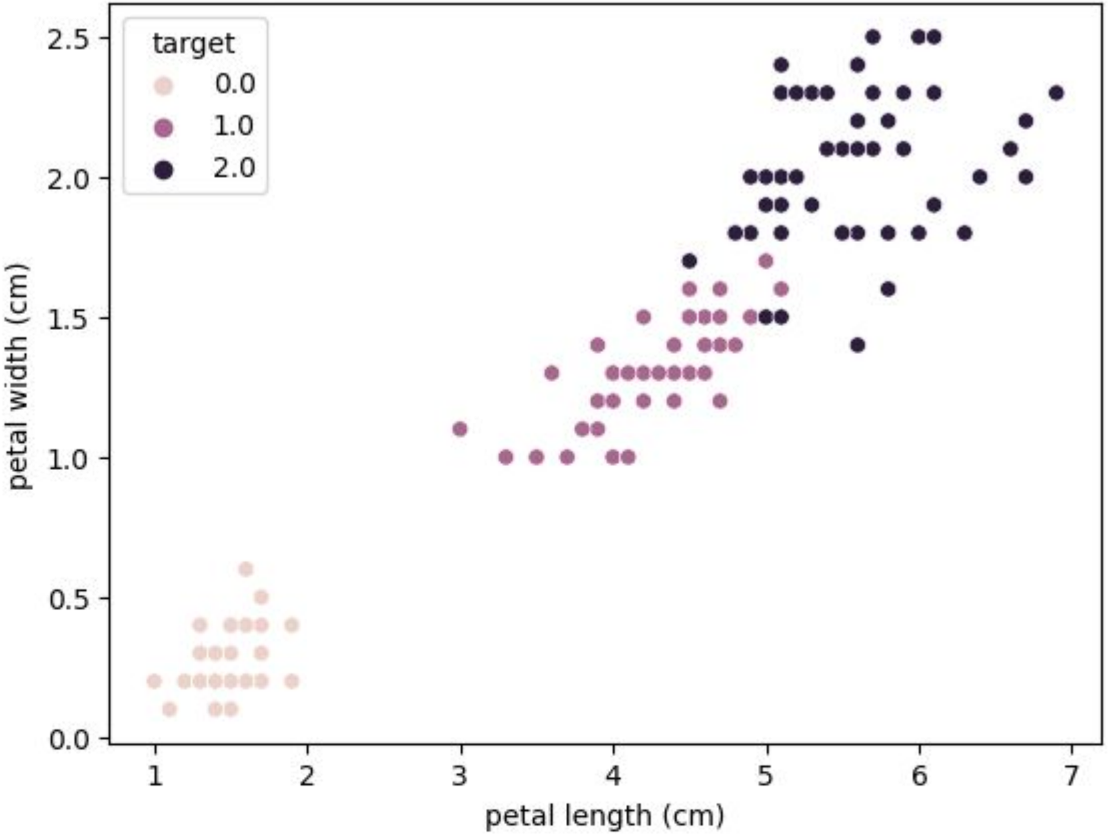
The Iris Dataset



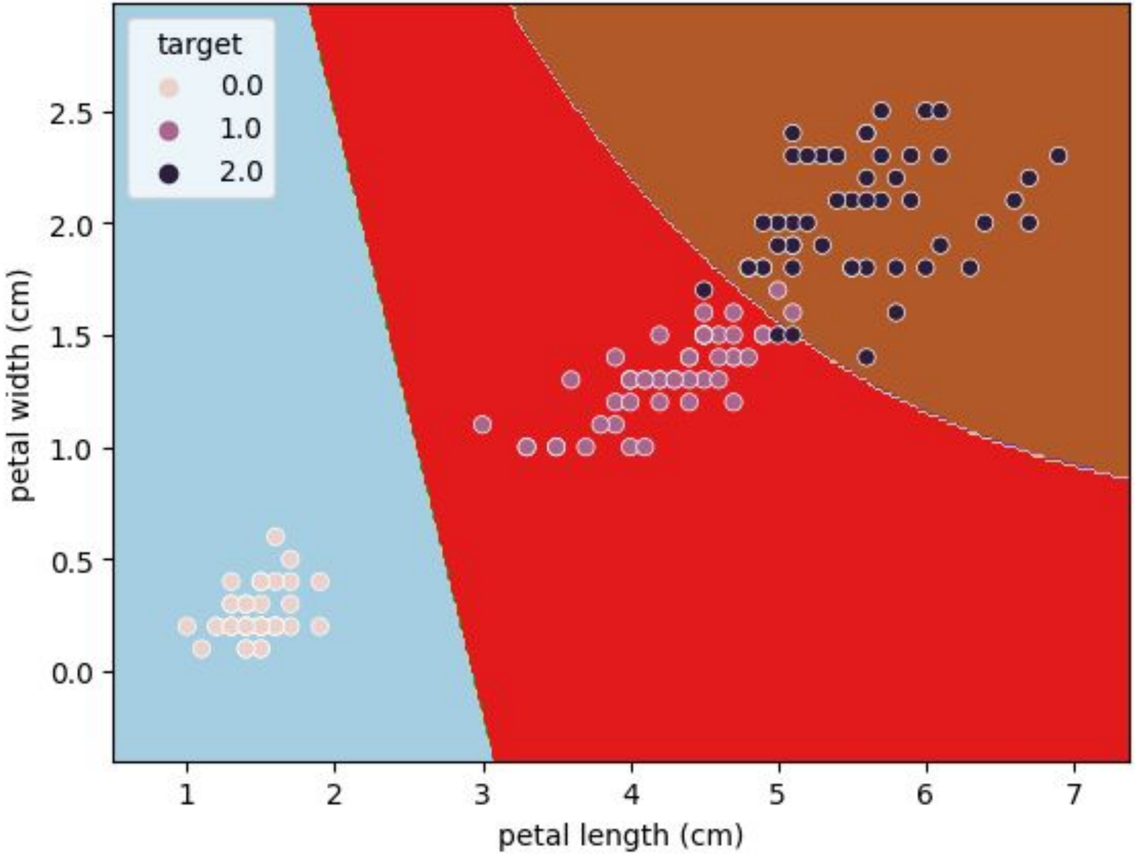
f(x) = y

Model type: Classification, Regression, Clustering

Complex Data

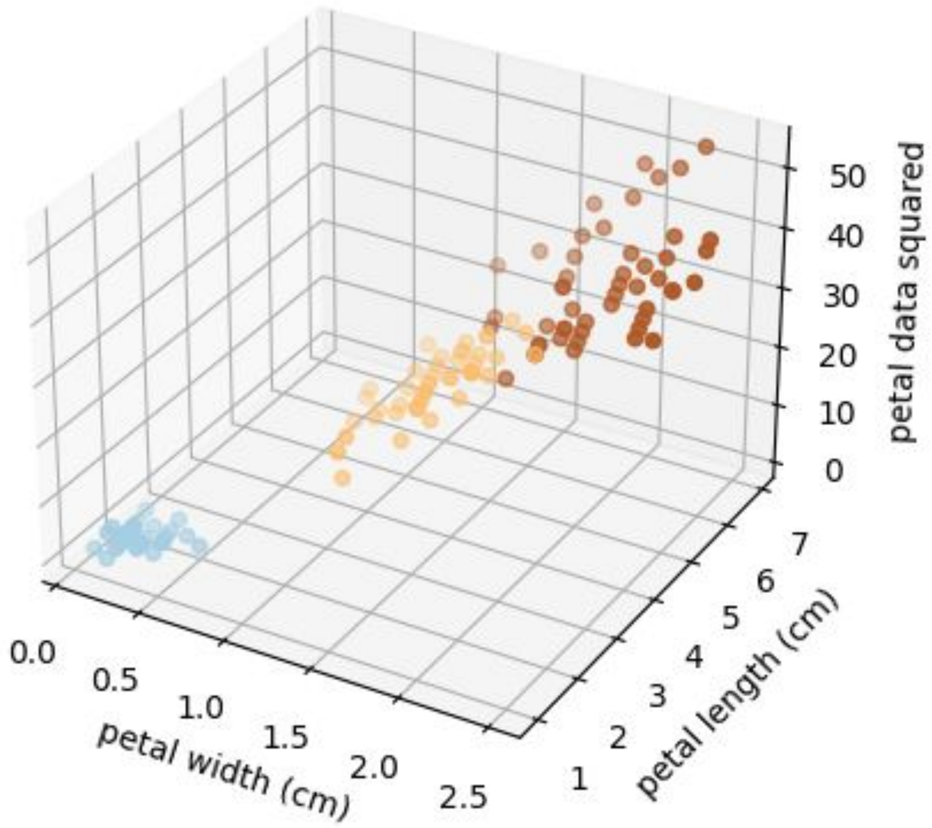


Using an SVM on Complex Data



Use a kernel to increase dimensionality

Here we square our values



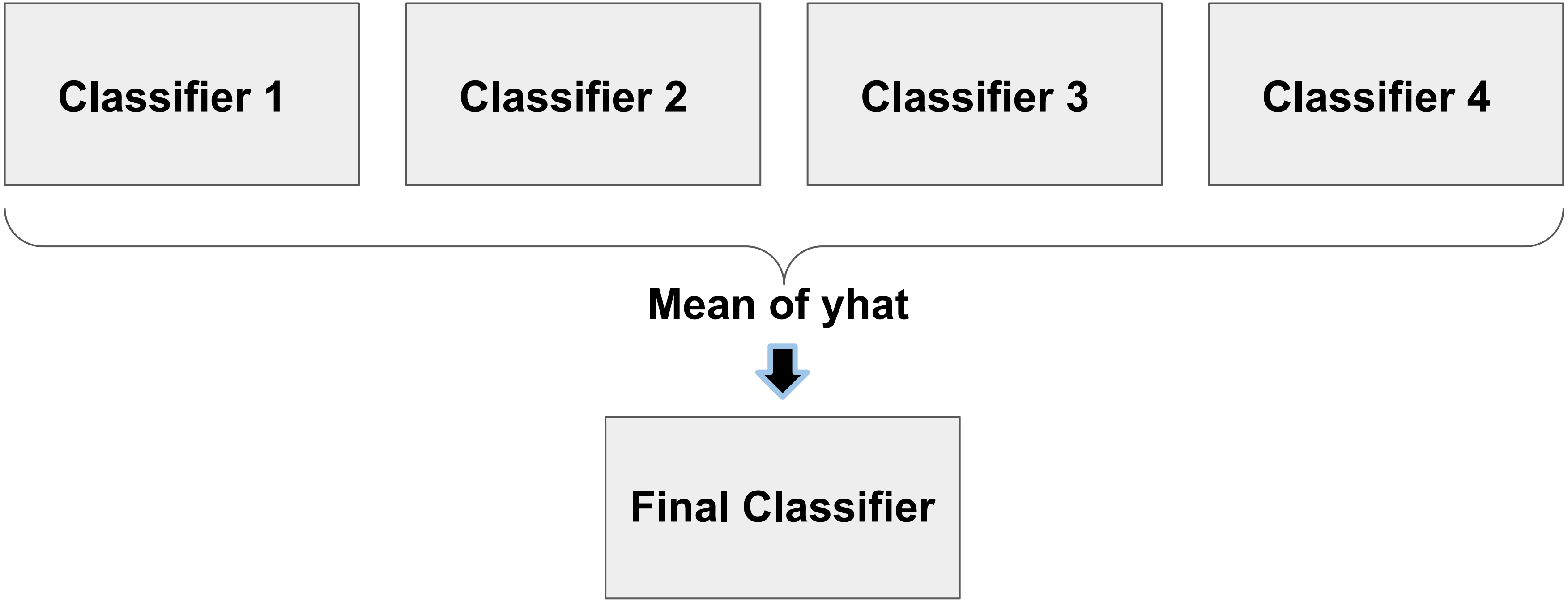
What is an Ensemble?

A mishmash of multiple instruments?

https://theclassicalnovice.com/glossary/ensembles/

What is a Machine Learning Ensemble?

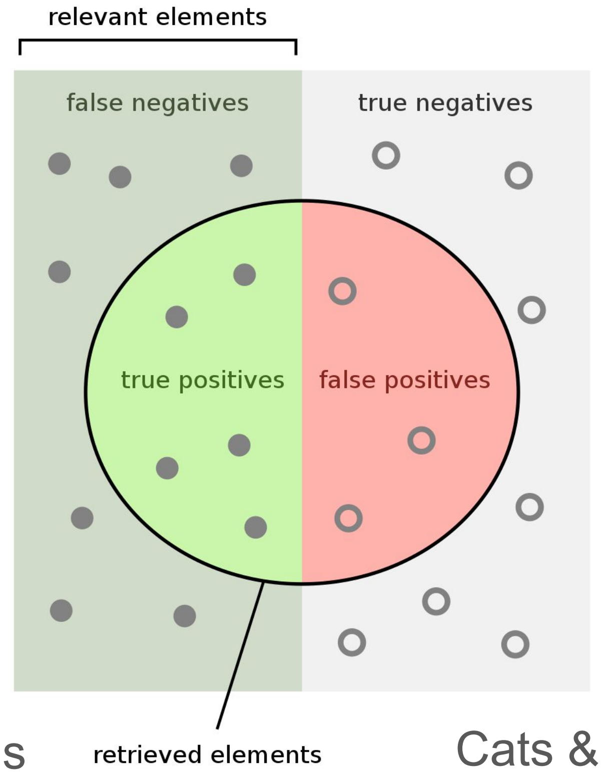
A mishmash of multiple learners



We will Cover Ensembles Next Week!

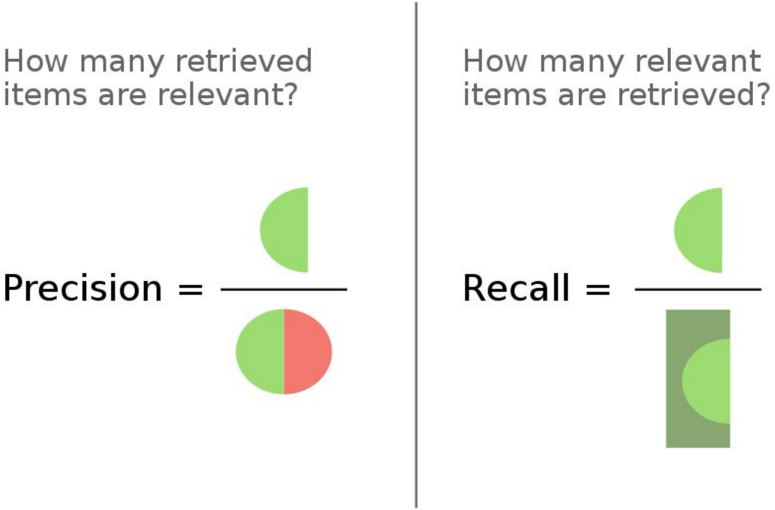
Jupyter Notebooks Time!

Machine Learning Evaluation Metrics (wiki)



Dogs

Cats & Donkeys



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Machine Learning Evaluation Metrics

TP,TN, FP, FN (True +, True -, False +, False -)

Precision and Recall

Receiver operating characteristic (ROC) curve and Area under curve (AUC)

Accuracy

F1 Score

- https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative
- https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall
- https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
- https://developers.google.com/machine-learning/crash-course/classification/accuracy
- https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

Brushing up on Covariance