

Source: https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

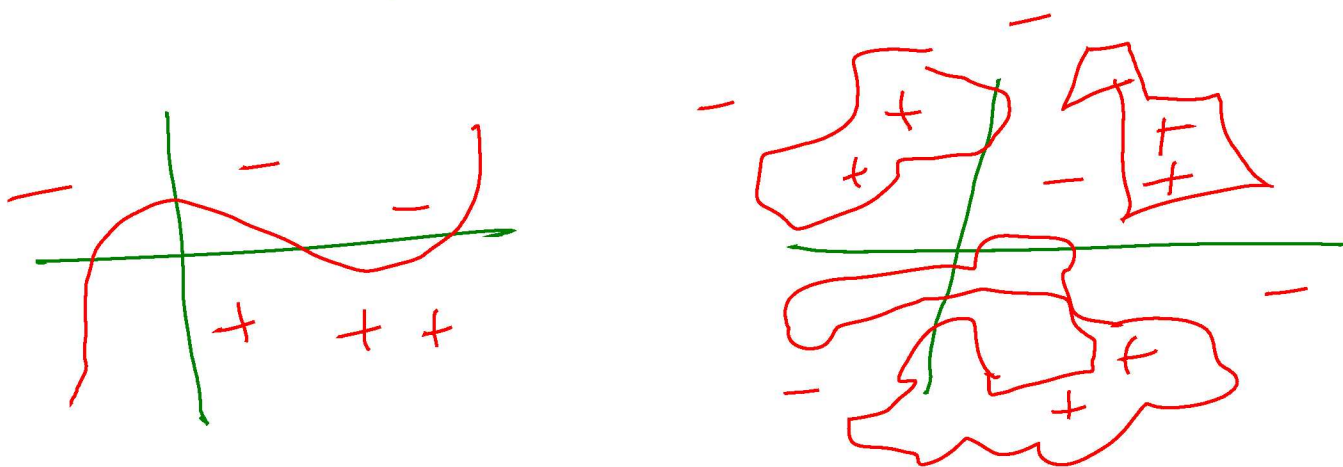
1 Pre-lecture Work

None. Get plenty of sleep and do well on all your midterms :)

2 Lecture

Problem 1. What is a decision boundary?

The "line" that separates + and - examples



Problem 2. What is a universal approximation theorem?

The model/hypothesis class is
able to achieve 0 training error
for all data sets

$$\underline{L_S}(h_S) = 0$$

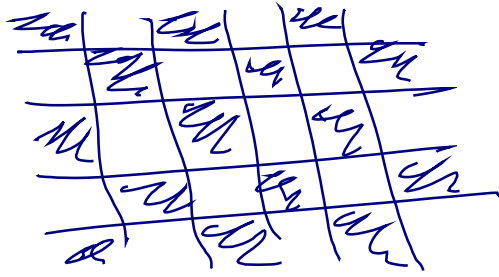
can't : $L_D(h_S) = 0$

There is some hyperparameter
such that as $\text{param} \rightarrow \infty$,
then $L_S(h_S) \rightarrow 0$

MLP: \mathcal{Q} ,

DT depth k

Ensembles T



Problem 3. Neural Networks.

1. Optional Videos:

(a) 3Blue1Brown on Neural Networks: https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQB0WTQDNU6R1_67000Dx_ZCJB-3pi

2. What is the hypothesis class of 1-layer neural networks?

$$\mathcal{H}_{\text{1-L}} = \left\{ x \mapsto \text{sign}(w^T \phi(x)) : w \in \mathbb{R}^{d_0} \right\}$$

Idea of NN/MLP is to automatically learn ϕ

$$\phi: \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1}$$

$$\phi = x \mapsto \sigma(Ax)$$

$$\sigma: \mathbb{R} \rightarrow \mathbb{R} \quad A: \mathbb{R}^{d_1 \times d_0}$$

also a param

3. What is the hypothesis class of n-layer neural networks?

hidden layer size

given a list of dim. $[d_1, \dots, d_n]$

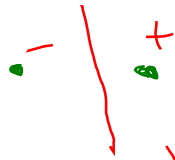
$$\mathcal{H}_{\text{MLP}} = \left\{ x \mapsto \text{sign} \left(w^T \sigma \left(A_n \sigma \left(A_{n-1} \sigma \left(\dots A_2 \sigma(A_1 x) \right) \right) \right) \right) \right\}$$

$$\bullet w \in \mathbb{R}^{d_n}, \quad A_i = \mathbb{R}^{d_i \times d_{i-1}}$$

4. What is the VC-dimension of neural networks? (Theorem 20.6)

$$= O(\log p) \quad p = \frac{d_0 d_1}{A_1} \frac{d_1 d_2}{A_2} \dots \frac{d_{n-1} d_n}{A_n}$$

$$X = \mathbb{R}^{d_0} \quad = d_0 \prod_{i=1}^n d_i^2$$

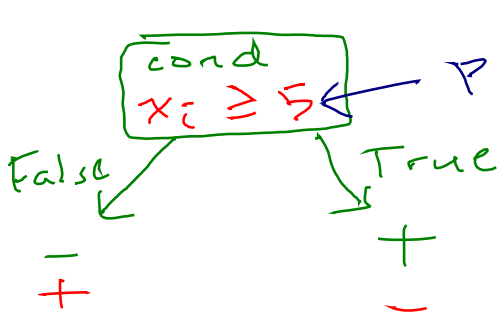


Problem 4. Decision Trees

1. Optional Videos:

- (a) StatQuest on decision trees: <https://www.youtube.com/watch?v=7VeUPuFGJHk>
- (b) StatQuest on regression trees: <https://www.youtube.com/watch?v=g9c66TUy1Z4>

2. What is the hypothesis class of decision stumps?



thresholds can't shatter 2 points

short tree, depth = 1
 $X = \mathbb{R}^d$, $Y = \{+1, -1\}$

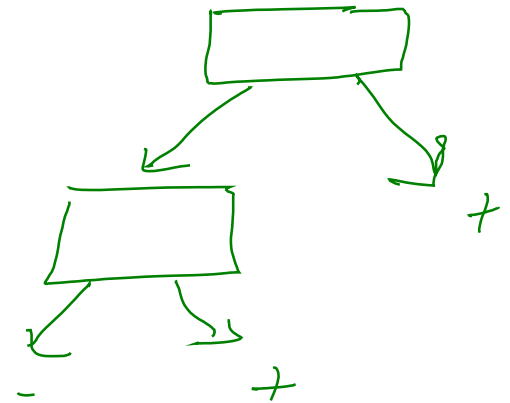
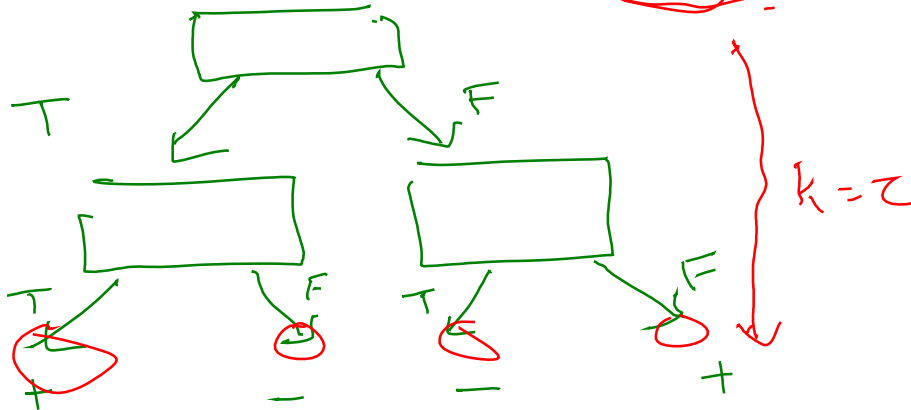
x_i is the i th coordinate

$$\mathcal{H}_{DS} = \left\{ x \mapsto \mathbb{I}[x_i \geq b] : i \in [d], s \in \{+1, -1\} \right\}$$

$\forall \text{dim} = 2$

$\{1, \dots, d\}$ $b \in \mathbb{R}$

3. What is the hypothesis class of depth k decision trees?



$$\# \text{ of leaves} \leq 2^k$$

4. What is the VC-dimension of depth k decision trees?

VCdim = number of leaf nodes

Ensemble + stumps = SOTA for vast majority of problems!

Problem 5. Ensemble Methods

1. Optional Videos:

- (a) StatQuest on random forests: https://www.youtube.com/watch?v=J4Wdy0Wc_xQ
- (b) StatQuest on AdaBoost: <https://www.youtube.com/watch?v=LsK-xG1cLYA>
- (c) StatQuest on XGBoost (4 videos): https://www.youtube.com/watch?v=OtD8wVaFm6E&list=PLblh5JK0oLUICTaGLRoHQDuF_7q2GfuJF&index=57
- (d) Alex Ihler on bagging: <https://www.youtube.com/watch?v=Rm6s6gmLTdg>

2. What is the hypothesis class of ensemble methods?

combine many votes from "weak" hypotheses into one "strong" hypothesis

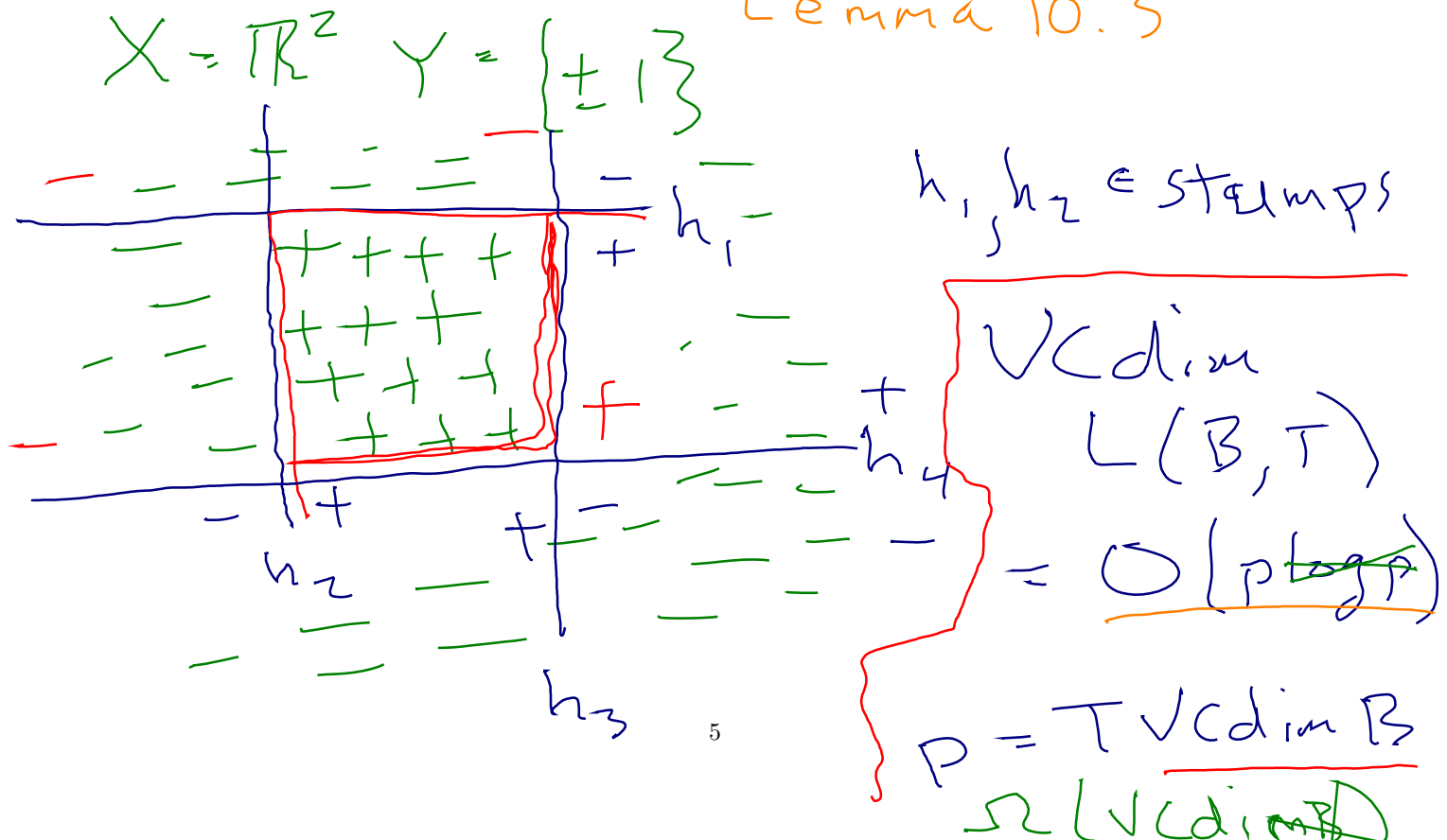
$$L(B, T) = \left\{ x \mapsto \text{sign} \left(\sum_{j=1}^T \omega_j h_j(x) \right) : \omega \in \mathbb{R}^T, h_j \in B \right\}$$

↑ another hyp class

↑ \mathbb{R}^T

3. What is the VC-dimension of ensemble methods?

Lemma 10.3

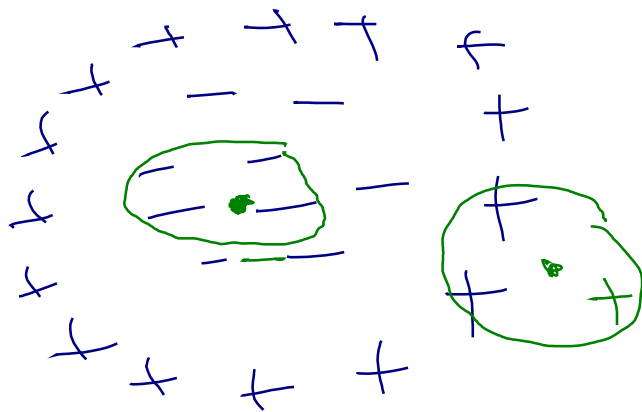
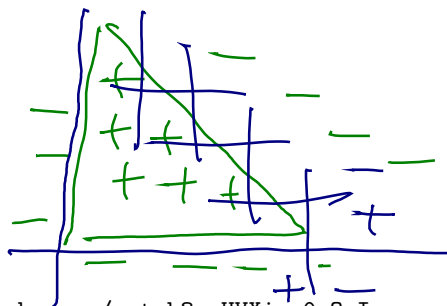


Problem 6. Nearest Neighbor

1. Optional Videos:

(a) StatQuest: <https://www.youtube.com/watch?v=HVXime0nQeI>

2. What is the k -nearest neighbor classification rule?



Take majority vote of k -nearest neighb.

3. What is the VC-dimension of k -nearest neighbor?

$\infty \Rightarrow$ not (agnostic) PAC learnable

4. Nearest neighbor can still be effective in practice. Why?