Final Exam Review

Xi Chen

- For a neural network, which one of these structural assumptions is the one that most affects the trade-off between underfitting (i.e. a high bias model) and overfitting (i.e. a high variance model):
  - (a): The number of hidden nodes
  - (b): The learning rate
  - (c): The initial choice of weights
  - (d): The use of a constant-term unit input

- For Kernel Regression, which one of these structural assumptions is the one that most affects the trade-off between underfitting and overfitting:
  - (a) Whether kernel is Gaussian versus triangular vs. box-shaped
  - (b) Whether we use Euclidian versus L1 versus L∞ metrics
  - (c) The kernel width
  - (d) The maximum height of the kernel function

VC-Dimensions

- T/F: If there exists a set of k instances cannot be shattered by H, then VC(H)$<k$

- Recall that for binary classification, shattered means for *any* label scheme of S , they can be correctly classified by a h$\in$H.
- To show VC(H)$\geq k$, there exists a set of instances can be shattered
- To show VC(H)$<k$: any set of k instances cannot be shattered
- (F)

VC-Dimensions

- T/F: If two hypothesis classes H_1 and H_2 satisfy $H_1 \subseteq H_2$, then $VC(H_1) \leq VC(H_2)$.
- (T)

- If three hypothesis classes H_1 , H_2, H_3 satisfy $H_1 = H_2 \cup H_3$, then $VC(H_1) \leq VC(H_2) + VC(H_3)$
- (F)

VC-Dimension Examples

$$H = \{h_\alpha | 0 \leq \alpha \leq 1, h_\alpha(x) = 1 \text{ iff } x \geq \alpha \text{ otherwise } h_\alpha(x) = 0\}.$$

VC(H)=1

$$H = \{h_{\mathbf{w}} | h_{\mathbf{w}} = \theta(w_0 + w_1 x_1 + w_2 x_2) \text{ where } \theta(z) = 1 \text{ iff } z \geq 0 \text{ otherwise } \theta_z = 0\}.$$

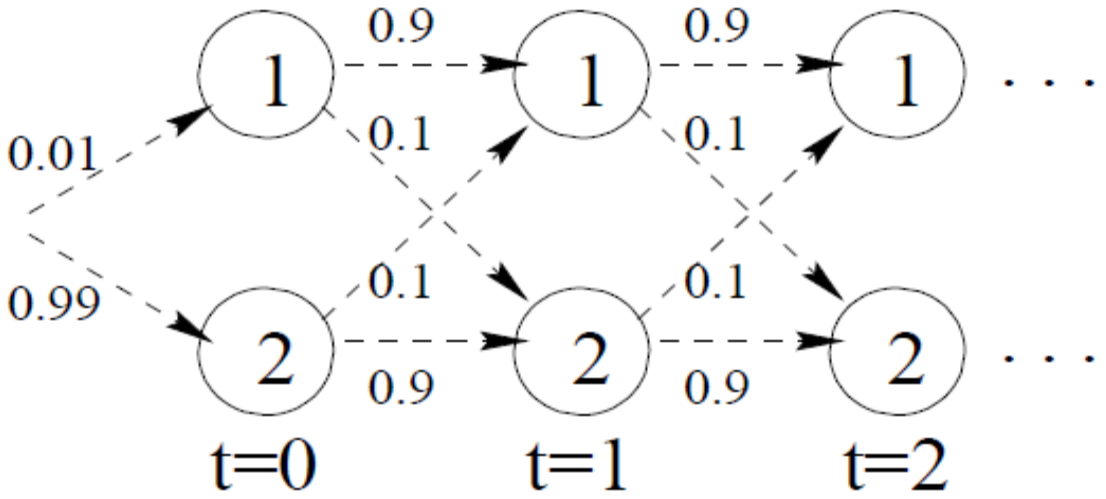
VC(H)=3

H is the set of all circles in 2D plane. Points inside the circles are classified as 1 otherwise 0.

VC(H)=3

HMM

$$P(x = heads|s = 1) = 0.51$$
$$P(x = heads|s = 2) = 0.49$$
$$P(x = tails|s = 1) = 0.49$$
$$P(x = tails|s = 2) = 0.51$$



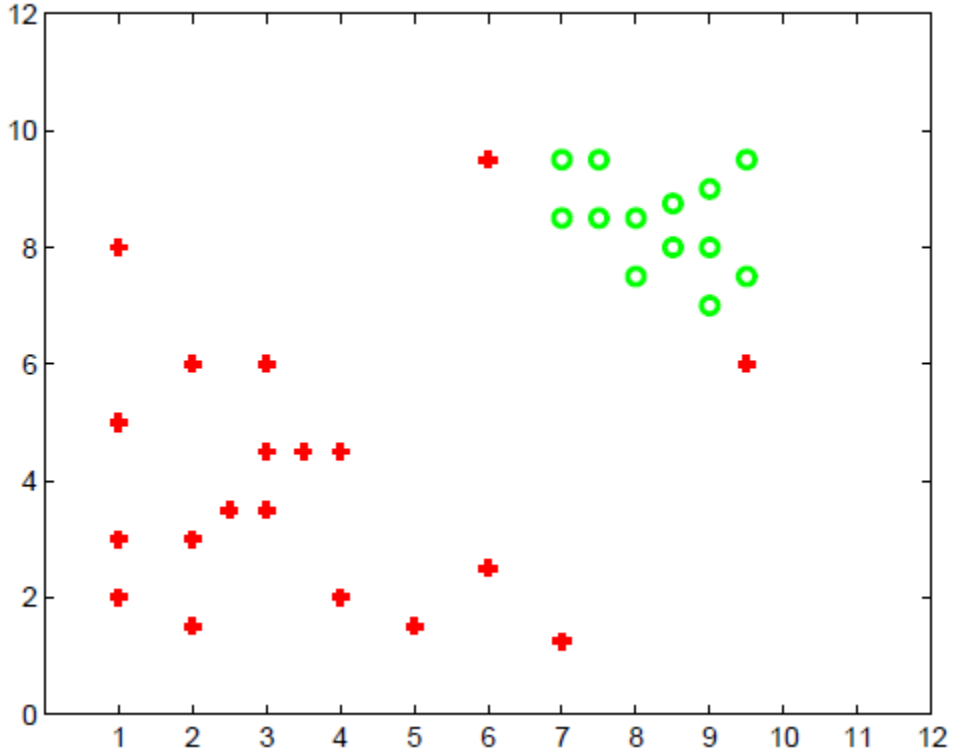
Q1: The sequence of observations is heads, heads, heads. What is the most likely state sequence given these three observations?

Answer: 2,2,2

Q2: What happens to the most likely state sequence if we observe a long sequence of all heads (e.g., 10^6 heads in a row)?

Answer: 2,1,1,1,1,...

SVM with Quadratic Kernel

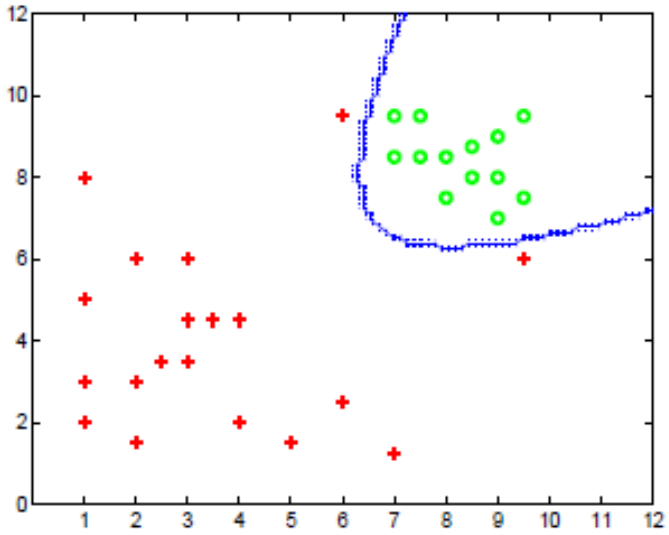


Decision Boundary:

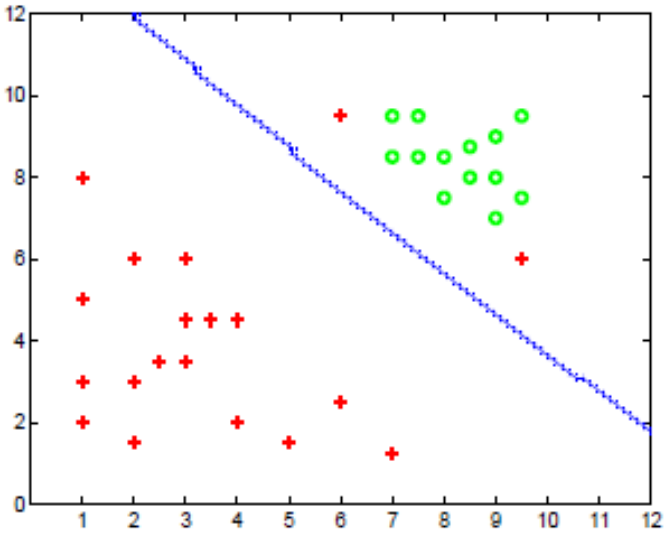
(a) $C \to \infty$
(b) $C=0$

(c) Which one is high-bias and low variance
(d) Which one is high-variance and low bias

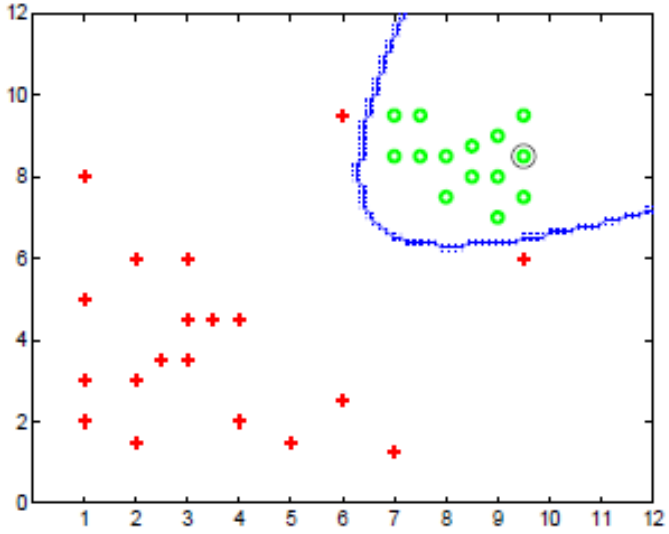
For $C \to \infty$
(e): Add points do not change the decision boundary
(f) : Add points which dramatically change the decision boundary



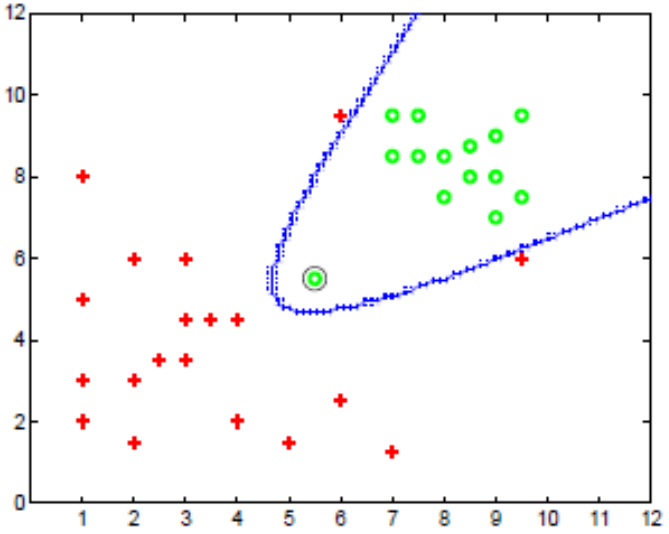
(a) Part 1



(b) Part 2



(c) Part 4



(d) Part 5