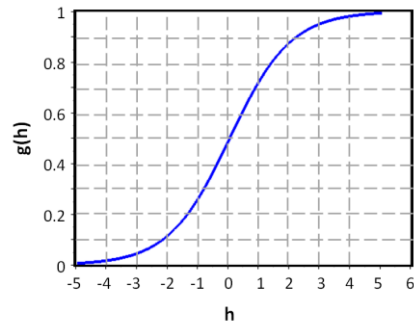
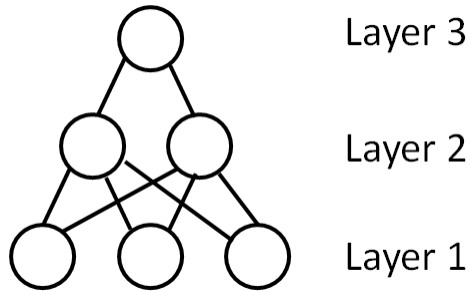


Final practice part 2

1. Consider the neural network below.



The initial weights are:

Layer 1:	$w_{1,1}^1 = -10$	$w_{1,2}^1 = 0$	$w_{1,3}^1 = -5$	$w_{1,4}^1 = 10$	$b_1^3 = 4$	Unit 1
	$w_{2,1}^1 = 20$	$w_{2,2}^1 = 0$	$w_{2,3}^1 = 10$	$w_{2,4}^1 = -5$	$b_1^3 = 4$	Unit 2
	$w_{3,1}^1 = 0$	$w_{3,2}^1 = -10$	$w_{3,3}^1 = 0$	$w_{3,4}^1 = 20$	$b_1^3 = 4$	Unit 3
Layer 2:	$w_{1,1}^2 = 5$	$w_{1,2}^2 = 10$	$w_{1,3}^2 = 0$	$b_1^3 = -2$		Unit 1
	$w_{2,1}^2 = 0$	$w_{2,2}^2 = -10$	$w_{2,3}^2 = 15$	$b_1^3 = -2$		Unit 2
Layer 3:	$w_{1,1}^3 = 10$	$w_{1,2}^3 = -20$	$b_1^3 = 5$			

Compute the output given the following inputs:

(a) Compute r_1^1, r_2^1, r_3^1 . Given the inputs: $x_1=5$ $x_2=-10$ $x_3=10$ $x_4=0$

(b) Compute r_1^3 . Given the lower-layer outputs: $r_1^2=0.1$, $r_2^2=0.6$

(c) Compute r_2^2 . Given the lower-layer outputs: $r_1^1=0.1$, $r_2^1=0.3$, $r_3^1=0.6$

Compute the change in the specified weight based on the following input/outputs. In each case, presume the starting weight is as specified in the original list above. Assume $\epsilon = 1$

(d) Compute $\Delta w_{1,2}^3$. Given the layer 2 rates: $r_1^2=0.2$ and $r_2^2=0.8$; layer 3 rates: $r_1^3=0.1$; the desired output from r_1^3 is 1.0

(e) Compute $\Delta w_{1,2}^1$. Given the features: $x_1=10$, $x_2=-5$, $x_3=0$, $x_4=15$; $r_1^1=0.5$, $r_2^1=0.2$, $r_3^1=0.8$; delta values: $\delta_1^2 = -0.005$, $\delta_2^2 = 0.01$

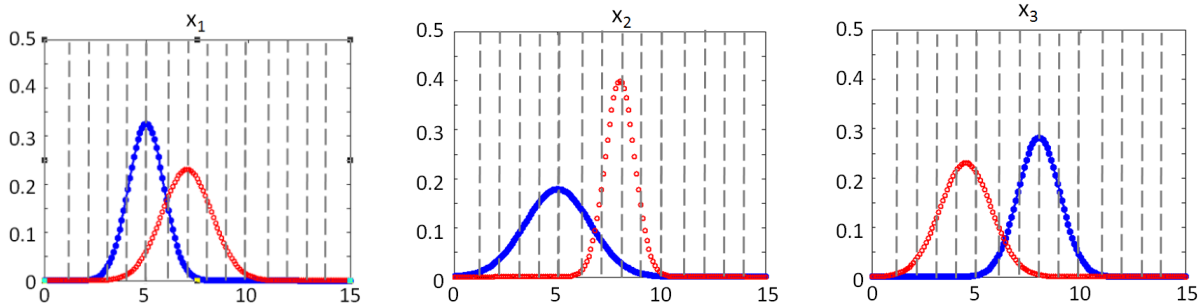
2. For each of the following functions $f(x; h)$, compute the value of h that will maximize $f(x; h)$, assuming each function has a single maximum and no minimum.

(a) $f_1(\mathbf{x}; h) = \sum_i (-h^2 - 10hx_i + 12x_i^2)$

(b) $f_2(x; h) = e^{-(h^3+x^2)} = \exp(-(h^2 + x^2))$

(c) $f_3(\mathbf{x}; h) = \prod_i 3h^{(x^i)}$

3. Consider the following Gaussian likelihoods for features x_1 , x_2 , and x_3 given **class = 1** (blue curves) or **class = 0** (red curves).



i. We wish to multiply these likelihoods together to compute $P(\mathbf{x} | y)$. Which type of classification is this:

- (a) Naïve Bayes Max-Posterior classification
- (b) Non-Naïve Bayes Max-Likelihood classification
- (c) Naïve Bayes Max-Posterior classification
- (d) Naïve Bayes Max-Likelihood classification
- (e) Support Vector Machine classification

ii. For the feature values below, which class is more probable (based on $P(\mathbf{x} | y)$ calculated from the plots above)?

(a) $x_1=5$ $x_2=7$ $x_3=6$

(b) $x_1=8$ $x_2=8$ $x_3=6$

iii. Which class is more probable if we also incorporate the following prior:

$$P(y=0) = 0.1 \quad P(y=1) = 0.9$$

to compute $P(y|\mathbf{x})$?

(a) $x_1=4$ $x_2=5$ $x_3=9$

(b) $x_1=6$ $x_2=7$ $x_3=7$

iv. Provide a prior that would make class 1 more probable if the \mathbf{x} values are:

$x_1=6$ $x_2=8$ $x_3=6$

4. Using each of the following kernel functions, compute the result of $K(\mathbf{x}^1, \mathbf{x}^2)$, for the specified input vectors.

$$K(\mathbf{c}, \mathbf{d}) = 2^{-(\mathbf{c}^T \mathbf{d} + 2)}$$

(a) $\mathbf{c} = \begin{bmatrix} 4 \\ 0 \\ -2 \end{bmatrix}$ $\mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 1.5 \end{bmatrix}$

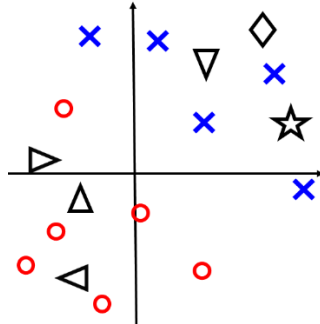
(b) $\mathbf{c} = \begin{bmatrix} 1 \\ 0.5 \\ -2 \end{bmatrix}$ $\mathbf{d} = \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}$

$$K(\mathbf{c}, \mathbf{d}) = (\mathbf{c}^T \mathbf{d} - 4)^2 + 10\mathbf{c}^T \mathbf{d}$$

(c) $\mathbf{c} = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}$ $\mathbf{d} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$

(d) $\mathbf{c} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$ $\mathbf{d} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$

5. Consider the following training data. Red circles are class 0, blue x's are class 1, and all other shapes (triangles, stars, diamonds) are data points with known feature values but unknown labels.



Using the EM approach for learning, and assuming that we use a linear logistic classifier, how will the black triangles, diamonds, and star data points be used for learning? In the first round of EM, what y value do you expect each data point to be assigned, or no value at all?

6. Consider the classification problem with the following features and classes.

Class Person-type: Teenager, YoungProfessional, Adult, SeniorCitizen

Features:

Daily-time-online: 1-2 hours, 3-4 hours, 5-8 hours

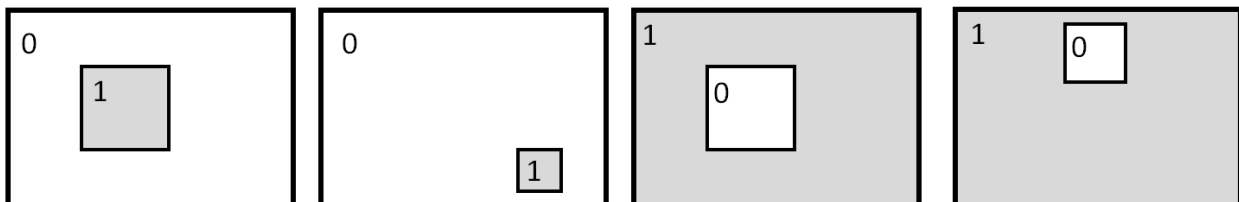
Number-of-online-friends: 0-10, 10-50, 50-200, 200-1000

Favored content: News, SocialPosts, Education, Entertainment

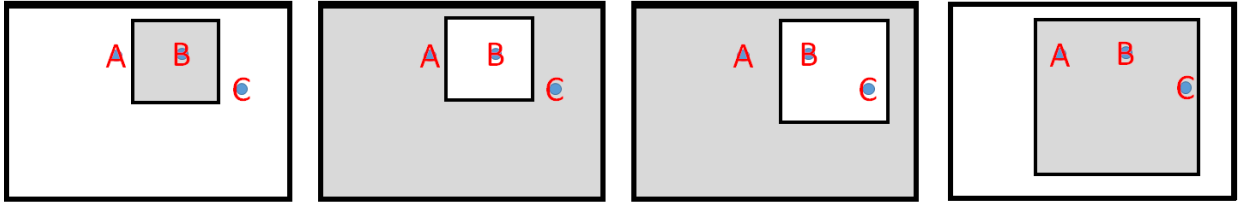
Money-spent-online: None, \$1-\$50, \$50-\$100, \$100-\$500, \$500+

(a) How many parameters given Naïve Bayes a posteriori classification?

Consider a classifier hypothesis set of squares. A single hypothesis h is a square with a fixed size and location. Four example hypotheses are shown.



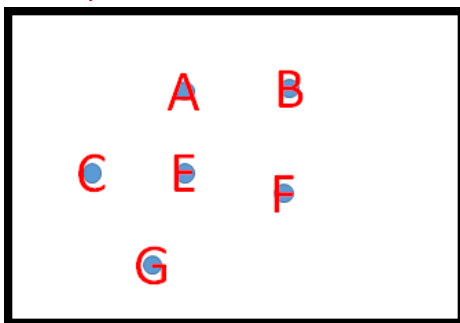
And here is examples of h that will help shatter a set of three data points.



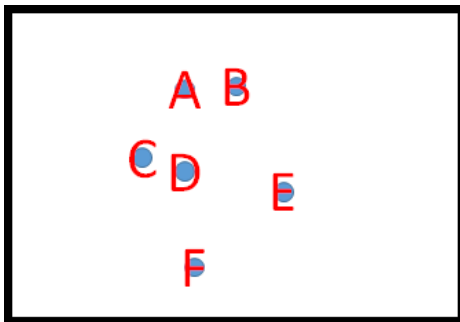
For each data set:

- Pick four points and list a dichotomy that is not possible with four of the points provided.

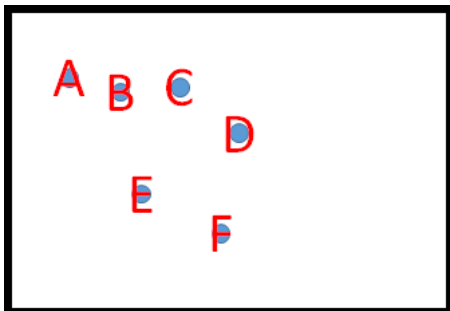
Example 1:



Example 2:



Example 3:



What is the VC dimension for a cube classifier, where each h is a cube of some fixed size and fixed location (x_1, x_2, x_3) , where class 1 is assigned either to inside or outside the cube, and class 0 is assigned to the other region. We assume you can select points from anywhere in the 3D feature space.

You have data with three features and wish to use a decision tree to classify your data (should have been covered in your past data mining class). For example, a two layer decision tree may look like this:

```
if  $x_1 > 0.5$  then:  if  $x_2 < -1$  :     $y=1$ 
                    else:           $y=0$ 
else:              if  $x_3 > 3$ :     $y=0$ 
                    else:           $y=1$ 
```

(First true-false measurement leads to second true-false, leading to a final class label.) We assume you can select points from anywhere in the 3D feature space.

What is the VC dimension for a one-layer decision tree?

~~What is the VC dimension for a two-layer decision tree? – I AM WITHDRAWING THIS QUESTION; IT IS MORE CHALLENGING THAN INTENDED~~

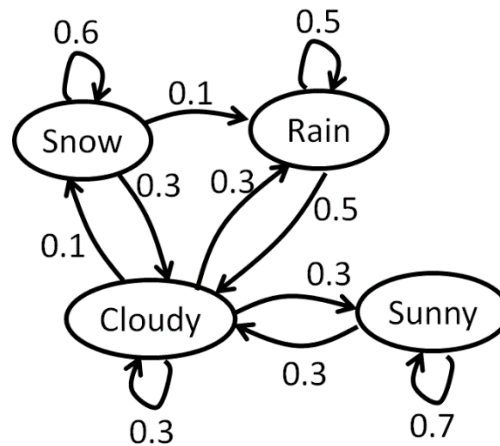
Consider the following HMM. It uses a thermometer to attempt to predict the weather.

We begin with the following estimate for our HMM parameters:

$$\Pi_{snow} = 0.2 \quad \Pi_{rain} = 0.3 \quad \Pi_{sunny} = 0.3 \quad \Pi_{cloudy} = 0.2$$

$\phi_{o,i}$:

	Cold	Mild	Hot
Snow	0.8	0.2	0
Rain	0.5	0.3	0.2
Sunny	0	0.3	0.7
Cloudy	0.2	0.7	0.1



(We COULD actually learn a Gaussian function for the temperature for each state. Here, we'll just do a discrete probability table.)

We receive a new sequence of temperatures and wish to update our HMM parameters.

Sequence:

Cold Cold Hot Mild Hot

Correct alpha values are in black. Made-up alpha values are in color parentheses. You will have to find the real values below. You can use the made-up value in calculating S_t values further below.

$\alpha_t(i)$

t:	1	2	3	4	5
Snow	?? (.11)	.08	0	.00011	0
Rain	0.15	?? (.04)	.0082	.0017	.00049
Sunny	?? (.08)	0	.0056	?? (.0033)	.0020
Cloudy	0.04	.027	?? (.0044)	.0053	.00030

Correct beta values are in black. Made-up beta values are in color parentheses. You will have to find the real values below. You can use the made-up value in calculating S_t values further below.

$\beta_t(i)$

t:	1	2	3	4
Snow	.0067	.0062	.13	.05
Rain	.0097	?? (.011)	.13	?? (.08)
Sunny	.0028	.087	?? (.11)	.52
Cloudy	.0062	.047	.121	?? (.11)

Find the missing values in the tables above.

What are the values:

$S_2(\text{cloudy})$

$S_3(\text{snow,sunny})$

$S_1(\text{rain})$

Now let us presume the following S values (these are made-up values):

$S_t(i)$

t	1	2	3	4	5
Snow	0.3	0.3	0.1	0.2	0.1
Rain	0.5	0.4	0.3	0.3	0.2
Sunny	0.1	0.1	0.3	0.1	0.4
Cloudy	0.1	0.2	0.3	0.4	0.3

$S_t(i,j)$

t	1	2	3	4
Rain, Cloudy	.1	.4	.3	.2
Sunny, Rain	0	0	0	0

Π_{rain}

Π_{cloudy}

$A_{rain,cloudy}$

$A_{sunny,rain}$

$\phi_{hot,rain}$

$\phi_{mild,sunny}$

Haven't had time to include CNN and Gauss mixture questions, but know:

- What is convolution?
- What is the max spatial pooling?
- What is normalization?
- How is sound converted to a picture for CNN?
- Rules for E and M step in Gauss mixture models, and understand how we got there.