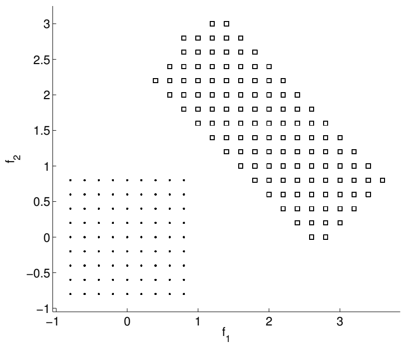
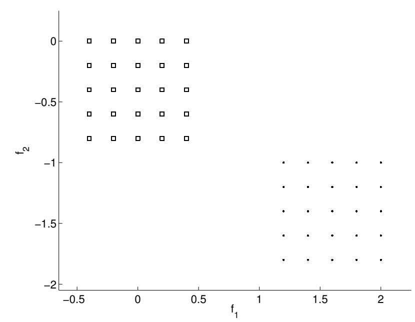
CS188: Exam Practice Session 10

Q1. Naïve Bayes Modeling Assumptions

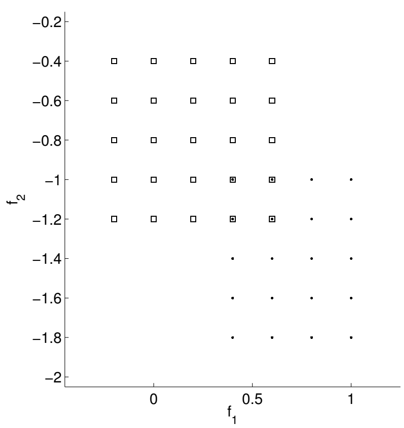
You are given points from 2 classes, shown as rectangles and dots. For each of the following sets of points, mark if they satisfy all the Naïve Bayes modelling assumptions, or they do not satisfy all the Naïve Bayes modelling assumptions. Note that in (c), 4 rectangles overlap with 4 dots.



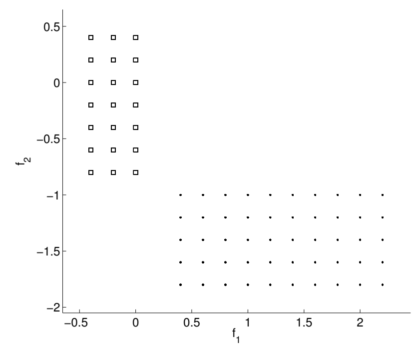
(a) ◯ Satisfies ◯ Does not Satisfy



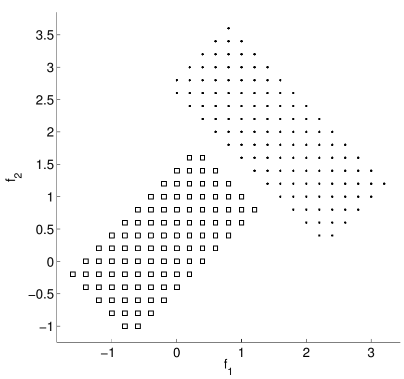
(b) ◯ Satisfies ◯ Does not Satisfy



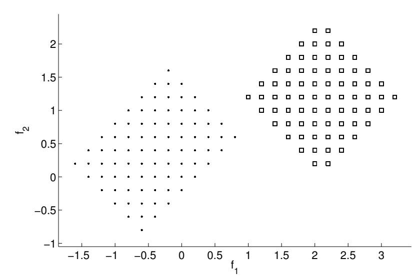
(c) ◯ Satisfies ◯ Does not Satisfy



(d) ◯ Satisfies ◯ Does not Satisfy



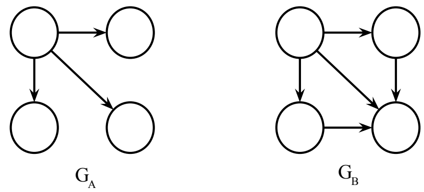
(e) ◯ Satisfies ◯ Does not Satisfy



(f) ◯ Satisfies ◯ Does not Satisfy

Q2. Model Structure and Laplace Smoothing

We are estimating parameters for a Bayes' net with structure G_A and for a Bayes' net with structure G_B. To estimate the parameters we use Laplace smoothing with $k = 0$ (which is the same as maximum likelihood), $k = 5$, and $k = \infty$.



Let for a given Bayes' net BN the corresponding joint distribution over all variables in the Bayes' net be P_{BN} then the likelihood of the training data for the Bayes' net BN is given by

$$\prod_{x_i \in \text{Training Set}} P_{BN}(x_i)$$

Let $\mathcal{L}_A^0$ denote the likelihood of the training data for the Bayes' net with structure G_A and parameters learned with Laplace smoothing with $k = 0$.
Let $\mathcal{L}_A^5$ denote the likelihood of the training data for the Bayes' net with structure G_A and parameters learned with Laplace smoothing with $k = 5$.
Let $\mathcal{L}_A^\infty$ denote the likelihood of the training data for the Bayes' net with structure G_A and parameters learned with Laplace smoothing with $k = \infty$.
We similarly define $\mathcal{L}_B^0$, $\mathcal{L}_B^5$, $\mathcal{L}_B^\infty$ for structure G_B.

For each of the questions below, mark which one is the correct option.

(a) Consider $\mathcal{L}_A^0$ and $\mathcal{L}_A^5$

○ $\mathcal{L}_A^0 \leq \mathcal{L}_A^5$ ○ $\mathcal{L}_A^0 \geq \mathcal{L}_A^5$ ○ $\mathcal{L}_A^0 = \mathcal{L}_A^5$ ○ Insufficient information to determine the ordering.

(b) Consider $\mathcal{L}_A^5$ and $\mathcal{L}_A^\infty$

○ $\mathcal{L}_A^5 \leq \mathcal{L}_A^\infty$ ○ $\mathcal{L}_A^5 \geq \mathcal{L}_A^\infty$ ○ $\mathcal{L}_A^5 = \mathcal{L}_A^\infty$ ○ Insufficient information to determine the ordering.

(c) Consider $\mathcal{L}_B^0$ and $\mathcal{L}_B^\infty$

○ $\mathcal{L}_B^0 \leq \mathcal{L}_B^\infty$ ○ $\mathcal{L}_B^0 \geq \mathcal{L}_B^\infty$ ○ $\mathcal{L}_B^0 = \mathcal{L}_B^\infty$ ○ Insufficient information to determine the ordering.

(d) Consider $\mathcal{L}_A^0$ and $\mathcal{L}_B^0$

○ $\mathcal{L}_A^0 \leq \mathcal{L}_B^0$ ○ $\mathcal{L}_A^0 \geq \mathcal{L}_B^0$ ○ $\mathcal{L}_A^0 = \mathcal{L}_B^0$ ○ Insufficient information to determine the ordering.

(e) Consider $\mathcal{L}_A^\infty$ and $\mathcal{L}_B^\infty$

○ $\mathcal{L}_A^\infty \leq \mathcal{L}_B^\infty$ ○ $\mathcal{L}_A^\infty \geq \mathcal{L}_B^\infty$ ○ $\mathcal{L}_A^\infty = \mathcal{L}_B^\infty$ ○ Insufficient information to determine the ordering.

(f) Consider $\mathcal{L}_A^5$ and $\mathcal{L}_B^0$

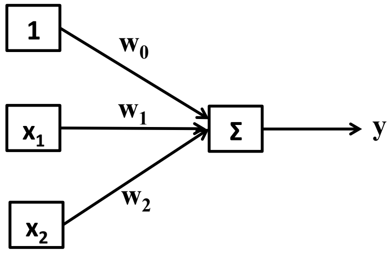
○ $\mathcal{L}_A^5 \leq \mathcal{L}_B^0$ ○ $\mathcal{L}_A^5 \geq \mathcal{L}_B^0$ ○ $\mathcal{L}_A^5 = \mathcal{L}_B^0$ ○ Insufficient information to determine the ordering.

(g) Consider $\mathcal{L}_A^0$ and $\mathcal{L}_B^5$

○ $\mathcal{L}_A^0 \leq \mathcal{L}_B^5$ ○ $\mathcal{L}_A^0 \geq \mathcal{L}_B^5$ ○ $\mathcal{L}_A^0 = \mathcal{L}_B^5$ ○ Insufficient information to determine the ordering.

Q3. Perceptron

(a) Consider the following perceptron, for which the inputs are the always 1 feature and two binary features $x_1 \in \{0, 1\}$ and $x_2 \in \{0, 1\}$. The output $y \in \{0, 1\}$.



$$y = \begin{cases} 1 & \text{if } (w_0 + w_1 \cdot x_1 + w_2 \cdot x_2) > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) Which one(s) of the following choices for the weight vector $[w_0\ w_1\ w_2]$ can classify y as $y = (x_1 \text{ XOR } x_2)$? XOR is the logical exclusive or operation, which equals to zero when x_1 equals to x_2 and equals to one when x_1 is different from x_2.

- ◯ [1 1 0]
- ◯ [-1.5 1 1]
- ◯ [-2 1 1.5]
- ◯ Any weights that satisfy $(-w_1 - w_2) < w_0 < \min(0, -w_1, -w_2)$.
- ◯ No weights can compute the XOR logical relation.

(ii) Which one(s) of the following choices for the weight vector $[w_0\ w_1\ w_2]$ can classify y as $y = (x_1 \text{ AND } x_2)$? Here AND refers to the logical AND operation.

- ◯ [1 1 0]
- ◯ [-1.5 1 1]
- ◯ [-2 1 1.5]
- ◯ Any weights that satisfy $(-w_1 - w_2) < w_0 < \min(0, -w_1, -w_2)$.
- ◯ No weights can compute the logical AND relation.

(b) Consider a multiclass perceptron with initial weights $w_A = [1\ 0\ 0]$, $w_B = [0\ 1\ 0]$ and $w_C = [\ 0\ 0\ 1]$. For prediction, if there is a tie, A is chosen over B over C. The following table gives a sequence of three training examples to be incorporated. When incorporating the second training example, start from the weights obtained from having incorporated the first training example. Similarly, when incorporating the third training example, start from the weights obtained from having incorporated the first training example and the second training example. Fill in the resulting weights in each row.

feature vector	label	w_A	w_B	w_C
		[1 0 0]	[0 1 0]	[0 0 1]
[1 -2 3]	A			
[1 1 -2]	B			
[1 -1 -4]	B			