Artificial Intelligence (521495A), Spring 2022
Exercise 4 : Machine learning
Solutions

This handout contains the example solutions for the problems 1-4.

Problem 1. Consider a dataset below where four binary features (i.e., input variables) are presented as f_1, f_2, f_3, and f_4, as well as corresponding binary class label y.

f_1	f_2	f_3	f_4	y
0	1	1	1	0
1	1	0	1	0
1	0	0	0	1
0	0	1	1	1
0	1	0	0	0
0	1	1	0	0

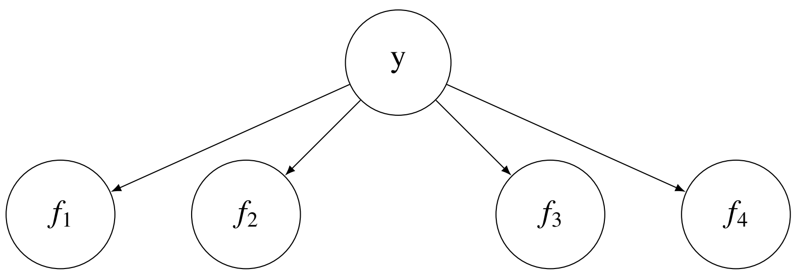
(a) Draw a Bayesian network for this problem (i.e., naive Bayes model).

(b) Calculate maximum likelihood estimate of $P(f_1 \mid y = 0)$.

(c) Calculate Laplace smoothed estimate of $P(f_3 \mid y = 1)$, with $k = 1$.

(d) Calculate Laplace smoothed estimate of $P(y = 1)$, with $k = 3$.

(a) The Bayesian network



Let's calculate these for $f_n = 1$. For $f_n = 0$ the value would be $1 - P$, respectively.

(b) $P(f_1 \mid y = 0) = \frac{c(f_1, y=0)}{c(y=0)} = \frac{1}{4}$

(c) $P(f_3 \mid y = 1) = \frac{c(f_3, y=1)+k}{c(y=1)+k|f_3|} = \frac{1+1}{2+1*2} = \frac{1}{2}$

(d) $P(y = 1) = \frac{c(y=1)+k}{n+k|y|} = \frac{2+3}{6+3*2} = \frac{5}{12}$

Problem 2. Football team manager has collected a labeled data set to be able to predict future game results. Target label of game result is y (+1: win, 0: tie, -1: lose) for the given input feature vector $\mathbf{x} = (h, w, t)$. There are three binary input features (1 or 0): h (home game or not), w (won the previous game or not), and t (team is in good shape or not).

h	w	t	y
1	1	1	+1
0	1	1	+1
1	0	1	0
0	1	0	0
0	0	1	-1
0	0	0	-1

(a) To predict future game results, manager is building a Naive Bayes (NB) classifier. Form and fill the model parameter tables for prior and conditional probabilities based on training data set above.

(b) Form a prediction equation for this NB classifier and calculate the prediction of the result of upcoming home game when team has win the previous game but has injured players (i.e., feature vector is $\mathbf{x} = (1, 1, 0)$). What will be the predicted result?

(c) Use Laplace smoothing with $k = 1$. What will be the predicted game result now?

(a) Model parameter tables:

P	$y = +1$	$y = 0$	$y = -1$
$P(\mathbf{y})$	2/6	2/6	2/6

P	$y = +1$	$y = 0$	$y = -1$
$P(h = 1 \mid \mathbf{y})$	1/2	1/2	0/2
$P(w = 1 \mid \mathbf{y})$	2/2	1/2	0/2
$P(t = 1 \mid \mathbf{y})$	2/2	1/2	1/2

(b) $y^* = \arg\max_y \{P(\mathbf{y})P(h \mid \mathbf{y})P(w \mid \mathbf{y})P(t \mid \mathbf{y})\}$

Predict $\mathbf{x} = (1, 1, 0)$:

$y = +1$: $P(y = +1)P(h = 1 \mid y = +1]P(w = 1 \mid y = +1)P(t = 0 \mid y = +1) = 2/6 * 1/2 * 2/2 * 0/2 = 0$

$y = 0$: $P(y = 0)P(h = 1 \mid y = 0]P(w = 1 \mid y = 0)P(t = 0 \mid y = 0) = 2/6 * 1/2 * 1/2 * 1/2 = 1/24$

$y = -1$: $P(y = -1)P(h = 1 \mid y = -1]P(w = 1 \mid y = -1)P(t = 0 \mid y = -1) = 2/6 * 0/2 * 0/2 * 1/2 = 0$

The prediction for $x = (1, 1, 0)$ is that the team will play tie (with 100% probability due to zero probabilities of other classes).

(c) Using Laplace smoothing with $k = 1$

Smoothed model parameter tables:

P	$y = +1$	$y = 0$	$y = -1$
$P(\mathbf{y})$	3/9	3/9	3/9

P	$y = +1$	$y = 0$	$y = -1$
$P(h = 1 \mid y)$	2/4	2/4	1/4
$P(w = 1 \mid y)$	3/4	2/4	1/4
$P(t = 1 \mid y)$	3/4	2/4	2/4

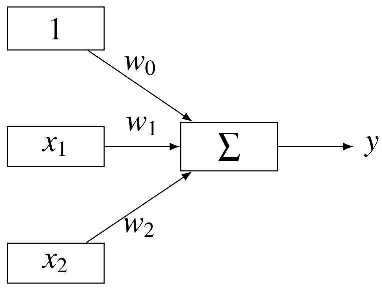
$y = +1$: $P(y = +1)P(h = 1 \mid y = +1]P(w = 1 \mid y = +1)P(t = 0 \mid y = +1) = 3/9 * 2/4 * 3/4 * 1/4 = 1/32$

$y = 0$: $P(y = 0)P(h = 1 \mid y = 0]P(w = 1 \mid y = 0)P(t = 0 \mid y = 0) = 3/9 * 2/4 * 2/4 * 2/4 = 1/24$

$y = -1$: $P(y = -1)P(h = 1 \mid y = -1]P(w = 1 \mid y = -1)P(t = 0 \mid y = -1) = 3/9 * 1/4 * 1/4 * 2/4 = 1/96$

Normalizing the probabilities gives $P(\mathbf{y} \mid \mathbf{x}) = (0.375, 0.50, 0.125)$. The prediction for $\mathbf{x} = (1, 1, 0)$ is that the game result is tie (with 50% probability). The winning class is still the same, but the confidence is now lower and other classes have non-zero probabilities due to smoothing. The probability of winning the game increases closer to winning class.

Problem 3. Consider the following perceptron model, for which the inputs are the constant 1 feature (bias) and two binary features $x_1 \in \{0, 1\}$ and $x_2 \in \{0, 1\}$. The output is $y \in \{0, 1\}$.



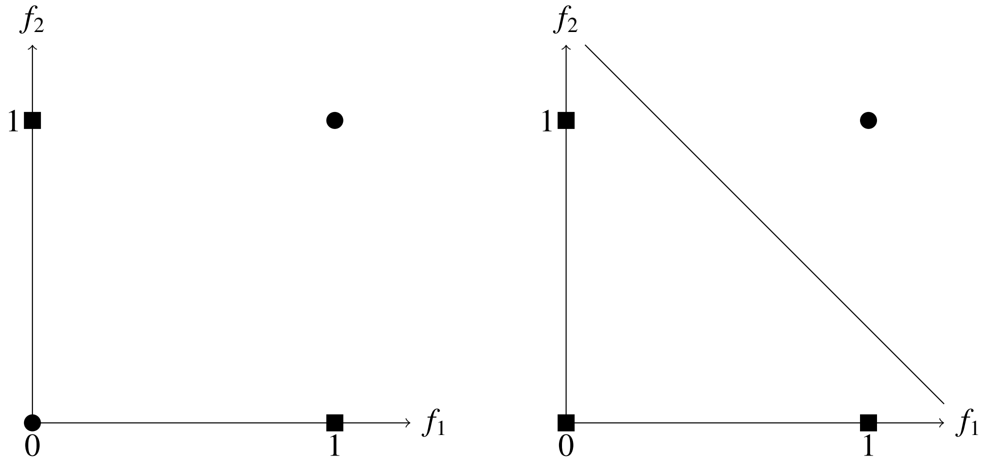
(a) Show visually if there is a proper solution (i.e., weight vector) for the logical XOR and AND operations using perceptron model. (XOR is the logical exclusive OR operation, which equals to zero when x_1 equals to x_2 and equals to one when x_1 is different from x_2)

(b) Show what conditions of weight vector satisfy the logical AND operation

(a) The truth table for the XOR and AND logical operations

x_1	x_2	XOR	AND
1	1	0	1
1	0	1	0
0	1	1	0
0	0	0	0

Feature spaces for the XOR and AND data



As seen from the figures, XOR data points cannot be separated using a single linear line (left), but AND data points can be separated with a single linear decision boundary (right).

(b) The derivation is as follows

In order to classify y as $y = x_1$ AND x_2 , we need to have $\begin{cases} w_0 + w_1 \cdot 1 + w_2 \cdot 1 > 0 \\ w_0 + w_1 \cdot 1 + w_2 \cdot 0 \leq 0 \\ w_0 + w_1 \cdot 0 + w_2 \cdot 1 \leq 0 \\ w_0 + w_1 \cdot 0 + w_2 \cdot 0 \leq 0 \end{cases}$.

It is equivalent to $\begin{cases} w_0 > -w_1 - w_2 \\ w_0 \leq -w_1 \\ w_0 \leq -w_2 \\ w_0 \leq 0 \end{cases}$, which is $(-w_1 - w_2) < w_0 \leq \min(0, -w_1, -w_2)$.

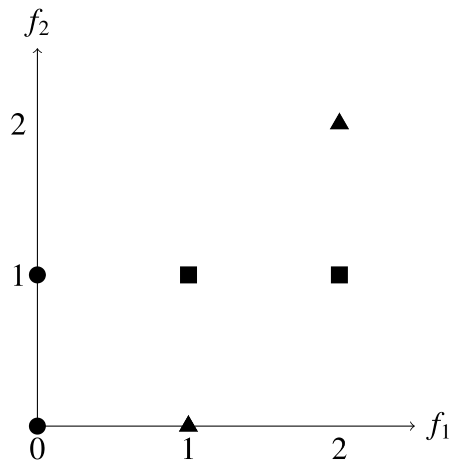
Problem 4. Consider a dataset below with two continuous features f_1 and f_2, taking values between 0-2 (i.e., $f_1 \in \mathbb{R}|0 \leq f_1 \leq 2$ and $f_2 \in \mathbb{R}|0 \leq f_2 \leq 2$), as well as corresponding discrete class label $y \in \{-1, 0, +1\}$.

f_1	f_2	y
0	0	-1
0	1	-1
1	1	0
2	1	0
1	0	+1
2	2	+1

(a) Show visually if classes are linearly separable.

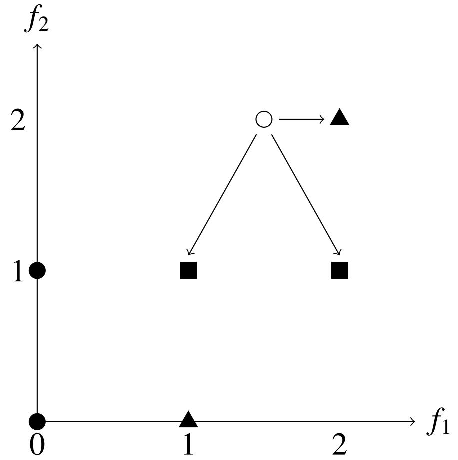
(b) An unknown example $\mathbf{x} = (f_1, f_2) = (1.5, 2)$ is given. What is the predicted class using k-nearest neighbor classifier with $k = 3$? (show visually)

(a) Feature space for given dataset



These three classes cannot be separated linearly, as seen from the illustration.

(b) Classify unknown example $x = (1.5, 2)$. Marked as non-filled circle in the figure below.



As illustrated above, using $k = 3$, the prediction is class $y = 0$, because it has majority of instances (i.e., filled rectangles) on a set of three closest data points. Although, the instance of $y = +1$ is the closest.