Problem Set 2

Problems 1-2 correspond to "The generative approach to classification"

Problem 1

1/1 point (graded)

Which of the following accurately describes the generative approach to classification, in the case where there are just two labels?

- Fit a model to the boundary between the two classes.
- Fit a probability distribution to each class separately.

Submit

Answers are displayed within the problem

Problem 2

1/1 point (graded)

In a generative model with k classes, the class probabilities are $\pi_1, \ldots, \pi_k$ (summing to 1) and the individual class distributions are $P_1(x), \ldots, P_k(x)$. In order to classify a new point x, we should pick the label j that maximizes which of the following quantities?

- π_j
- $P_j(x)$

$\pi_j + P_j(x)$

$\pi_j P_j(x)$

Submit

Answers are displayed within the problem

Problems 3-8 correspond to "Probability review I: probability spaces, events, conditioning"

Problem 3

3/3 points (graded)

What is the **size** of the **sample space** in each of the following experiments?

a) A fair coin is tossed.

2 **Answer:** 2

2

Answer

Correct: The possible outcomes are 0 and 1.

b) A fair die is rolled.

6 **Answer:** 6

6

Answer

Correct: The possible outcomes are 1,2,3,4,5,6.

c) A fair coin is tossed ten times in a row.

2^10 **Answer:** 1024

2^{10}

Answer
Correct:
For each of the coins, there are two possible outcomes. For all ten coins together, there are $2 \times 2 \times \cdots \times 2 = 1024$ outcomes.

? **Hint (1 of 1):** The sample space is the set of possible outcomes. How many possibilities are there in each case?

Next Hint

Submit

Answers are displayed within the problem

Problem 5

3/3 points (graded)
Two fair dice are rolled. What is the probability that:

a) Their sum is 10, given that the first roll is a 6?

1/6

Answer: 1/6

$\frac{1}{6}$

Answer
Correct: If the first roll is a 6, the second needs to be a 4, which happens with probability 1/6.

b) Their sum is 10, given that the first roll is an even number?

4/36

Answer: 1/9

$\frac{4}{36}$

Answer
Correct:
The probability that the sum is 10 *given that* the first roll is even is, by the basic conditioning formula, equal to Pr(sum is 10 AND first roll is even) divided by Pr(first roll is even). Let's compute these two separately. Pr(sum is 10 AND first roll is even) correspond to just two possible outcomes, (4,6) and (6,4); the probability that one of these occurs is 2/36 = 1/18. Meanwhile, Pr(first roll is even) is 1/2. Now divide.

c) They have the same value?

6/36

✔ **Answer:** 1/6

$\frac{6}{36}$

Answer

Correct: Whatever the first roll is, the probability that the second roll is exactly that number is 1/6.

Submit

Answers are displayed within the problem

Problem 6

1/1 point (graded)

A certain genetic disease occurs in 5% of men but just 1% of women. Let's say there are an equal number of men and women in the world. A person is picked at random and found to possess the disease. What is the probability, given this information, that the person is male?

5/6

✔ **Answer:** 5/6

$\frac{5}{6}$

Explanation

This is an application of Bayes' rule. Let's pick a person at random, and let D denote the event that they have the disease, and M the event that they are male. We want $Pr\left(M|D\right)$. By Bayes' rule,

$$Pr\left(M|D\right) = Pr\left(M\right) \times \frac{Pr(D|M)}{Pr(D)} = \frac{1}{2} \times \frac{0.05}{(1/2)\times 0.05+(1/2)\times 0.01} = \frac{5}{6}.$$

? **Hint (1 of 2):** You can use Bayes' rule for this. Next Hint

Submit

Answers are displayed within the problem

Problem 7

2/2 points (graded)

The TryMe smartphone company has three factories making its phones. They are all fairly unreliable: 10% of the phones from factory 1 are defective, 20% of the phones from factory 2 are defective, and 24% of the phones from factory 3 are defective. The factories do not produce the same numbers of phones: factory 1 produces $1/2$ of TryMe's phones, while factories 2 and 3 each produce $1/4$.

a) What is the probability that a TryMe phone chosen at random is defective?

4/25 ✔ **Answer:** 0.16

$\frac{4}{25}$

Answer

Correct:

For a phone chosen at random, let D denote the event that it is defective, F_1 that it comes from factory 1, F_2 that it comes from factory 2, and F_3 that it comes from factory 3. Then $Pr(D) = Pr(D \cap F_1) + Pr(D \cap F_2) + Pr(D \cap F_3)$. Applying the formula for conditional probability, we then have $Pr(D) = Pr(F_1)\, Pr(D|F_1) + Pr(F_2)\, Pr(D|F_2) + Pr(F_3)\, Pr(D|F_3)$. We have all the information we need for the right-hand side; plugging in, $Pr(D) = \frac{1}{2} \times 0.1 + \frac{1}{4} \times 0.2 + \frac{1}{4} \times 0.24 = 0.16$.

b) Given that a TryMe phone is defective, what is the probability that it came from factory 1?

5/16 ✔ **Answer:** 5/16

$\frac{5}{16}$

Answer

Correct: By Bayes' rule, $Pr(F_1|D) = Pr(F_1) \times \frac{Pr(D|F_1)}{Pr(D)} = \frac{1}{2} \times \frac{0.1}{0.16} = \frac{5}{16}$.

? **Hint (1 of 3):** For a phone chosen at random, let D denote the event that it is defective, F_1 that it comes from factory 1, F_2 that it comes from factory 2, and F_3 that it comes from factory 3. For part (a), we want $Pr(D)$. For part (b), we want $Pr(F_1|D)$.

Next Hint

Submit

Answers are displayed within the problem

Problem 8

1/1 point (graded)

Here are some statistics collected by a doctor about patients who walk into her office.

- 25% of the patients have the flu.
- Among patients with the flu, 75% have a fever.
- Among patients who don't have the flu, 50% have a fever.

A new person walks into the doctor's office and turns out to have a fever. What is the probability that he has the flu?

1/3

✔ **Answer:** 1/3

$\frac{1}{3}$

Explanation

By Bayes' rule,

$Pr\left(\text{flu}|\text{fever}\right) = Pr\left(\text{flu}\right) \times \frac{Pr(\text{fever}|\text{flu})}{Pr(\text{fever})} = 0.25 \times \frac{0.75}{Pr(\text{fever})}$.

We can compute the probability of having a fever by splitting it into two cases:

$Pr\left(\text{fever}\right) = Pr\left(\text{flu}\right) Pr\left(\text{fever}|\text{flu}\right) + Pr\left(\text{no flu}\right) Pr\left(\text{fever}|\text{no flu}\right) = 0.25 \times 0.75 + 0.75 \times 0.5.$

Putting this all together gives the answer.

Hint (1 of 2): By Bayes' rule,

$Pr\left(\text{flu}|\text{fever}\right) = Pr\left(\text{flu}\right) \times \frac{Pr(\text{fever}|\text{flu})}{Pr(\text{fever})}$

Next Hint

Submit

Answers are displayed within the problem

Problems 9-12 correspond to "Generative modeling in one dimension"

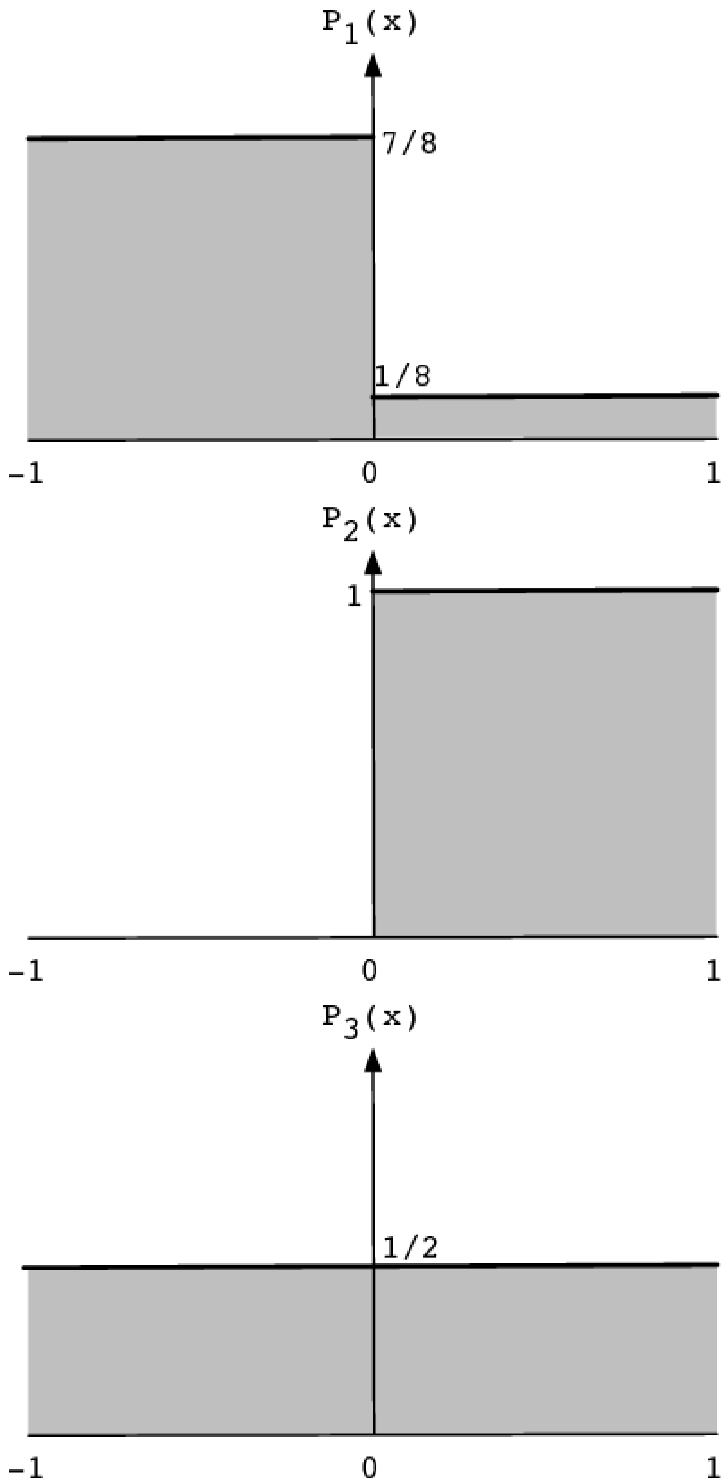
Problem 9

2/2 points (graded)

Suppose we have one-dimensional data points lying in $X = [-1, 1]$, that have associated labels in $Y = \{1, 2, 3\}$. The individual classes have weights

$$\pi_1 = \frac{1}{3}, \quad \pi_2 = \frac{1}{6}, \quad \pi_3 = \frac{1}{2}$$

and densities P_1, P_2, P_3 as shown below. (For instance, P_1 is the density of the points whose label is 1; in particular, this means that P_1 integrates to 1.)



Based on this information, what labels should be assigned to the following points?

a) $-1/2$

1

Answer: 1

1

b) $1/2$

3

Answer: 3

3

Explanation
In each case (for each of the two given values of x), we need to compute $\pi_1 P_1(x), \pi_2 P_2(x), \pi_3 P_3(x)$, and then select the label that has the largest value.

Submit

Answers are displayed within the problem

Problem 10

2/2 points (graded)
A set of 100 data points in $\mathbb{R}$ have mean of 20 and standard deviation of 10. We want to fit a Gaussian $N(\mu, \sigma^2)$ to this data. What μ and σ^2 should we pick?

a) μ =

20

Answer: 20

20

b) σ^2 =

100

Answer: 100

100

Explanation
μ is the mean, 20, and σ^2 is the variance, $10^2 = 100$.

Submit

Answers are displayed within the problem

Problem 11

1/1 point (graded)

A generative approach is used for a binary classification problem and it turns out that the resulting classifier predicts $+$ at $\mathbf{all}$ points x in the input space. What can we conclude for sure? Check all that apply.

- [ ] There are no $-$ points in the training set.
- [ ] The $+$ points are spread out over the space, while the $-$ points are concentrated in a small region.
- [x] There are fewer $-$ points than $+$ points in the training set.
- [ ] The density of $+$ points is greater than the density of $-$ points everywhere in the space.

Explanation

First option: a possible explanation, but it isn't necessarily the case.
Second option: In this case, we would expect to predict $-$ in that small region.
Third option: Yes. We are told that $\pi_+ P_+(x) > \pi_- P_-(x)$ for all x. At the same time, since probability distributions must sum to 1, there must be at least one point x at which $P_-(x) \geq P_+(x)$. In order for the prediction at that point to be $+$, we must have $\pi_+ > \pi_-$.
Last option: This is not mathematically possible, since the density of a probability distribution always has to sum/integrate to 1.

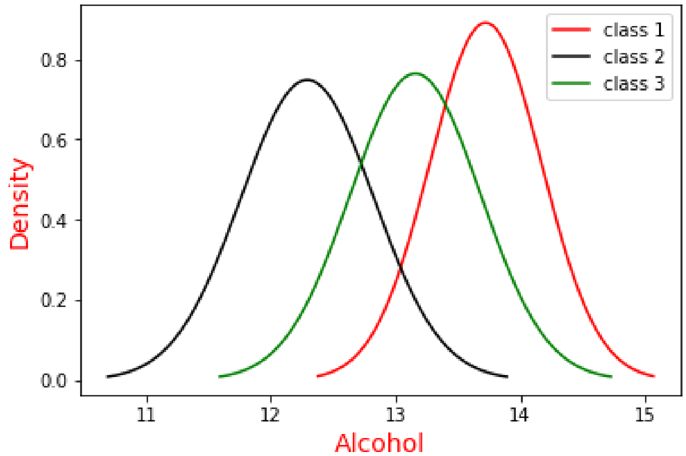
Submit

Answers are displayed within the problem

Problem 12

5/5 points (graded)

For the winery example from lecture, the densities obtained are reproduced here:



The class probabilities are $\pi_1 = 0.33, \pi_2 = 0.39, \pi_3 = 0.28$. What labels would be assigned to the following points?

a) 12.0

2

✔ **Answer:** 2

2

b) 12.5

2

✔ **Answer:** 2

2

c) 13.0

3

✔ **Answer:** 3

3

d) 13.5

1

✔ **Answer:** 1

1

e) 14.0

1

✔ **Answer:** 1

1

Explanation

This is a matter of eyeballing the densities and figuring out which of $\pi_1 P_1(x), \pi_2 P_2(x), \pi_3 P_3(x)$ is the largest in each case.

Submit

Answers are displayed within the problem

Problems 13-15 correspond to "Probability review II: random variables, expected value, and variance"

Problem 13

4/4 points (graded)

A fair die is rolled twice. Let X_1 and X_2 denote the outcomes, and define random variable X to be the minimum of X_1 and X_2.

a) How many possible values are there for X?

6

✔ **Answer:** 6

6

Answer

Correct: The minimum of the two rolls could be any number from 1 to 6.

b) What is the probability that $X = 1$?

11/36

✔ **Answer:** 11/36

$\frac{11}{36}$

Answer

Correct: This is the probability that at least one of the two rolls is a 1.

c) What is $E(X)$?

2.52

✔ **Answer:** 91/36

2.52

d) What is $\text{var}(X)$?

1.97

✔ **Answer:** 2555/1296

1.97

? **Hint (1 of 3):** Notice that the possible values of X are 1,2,3,4,5,6, but these aren't equally likely. You need to compute the distribution of X, that is, you need to figure out $Pr(X=1), Pr(X=2), Pr(X=3), Pr(X=4), Pr(X=5), Pr(X=6)$.

Next Hint

Submit

Answers are displayed within the problem

Problem 14

2/2 points (graded)

In a series of ten independent experiments, a random variable X takes on values

$$1, 1, 2, 5, 0, 1, 2, 2, 1, 1.$$

a) Give an estimate of $E(X)$.

8/5

✔ **Answer:** 1.6

$\frac{8}{5}$

b) Give an estimate of $\text{var}(X)$.

41/25

✔ **Answer:** 1.64

$\frac{41}{25}$

? **Hint (1 of 1):** Just compute the mean and variance assuming $Pr\left(X=0\right)=0.1, Pr\left(X=1\right)=0.5, Pr\left(X=2\right)=0.3, Pr\left(X=5\right)=0.1$
.

Next Hint

Submit

Answers are displayed within the problem

Problem 15

1/1 point (graded)
Which of the following random variables has $\mathbf{zero\ variance}$? Check all that apply.

- [ ] X takes on values -1 and 1 with equal probability.
- [x] X always takes on value 1.
- [ ] X is always equal to X^2.
- [x] X is always zero.

✔

Explanation
Zero variance implies that X always takes on the same value. For the third option, note that X could be either 0 or 1.

Submit

Answers are displayed within the problem

Problems 16-18 correspond to "Probability review III: modeling dependence"

Problem 16

4/4 points (graded)

In each of the following cases, say whether X and Y are dependent or independent.

a) Randomly pick a card from a pack of 52 cards. Define X to be 1 if the card is a Jack, and 0 otherwise. Define Y to be 1 if the card is a spade, and 0 otherwise.

- ○ dependent
- ● independent

✔

b) Randomly pick two cards from a pack of 52 cards. X is 1 if the first card is a spade, and 0 otherwise. Y is 1 if the second card is a spade, and 0 otherwise.

- ● dependent
- ○ independent

✔

c) Toss a coin ten times. X is the number of heads and Y is the number of tails.

- ● dependent
- ○ independent

✔

d) Roll a fair die. X is 1 if the outcome is even, and 0 otherwise. Y is 1 if the outcome is ≥ 3, and zero otherwise.

- ○ dependent
- ● independent

✔

Submit

Answers are displayed within the problem

Problem 17

2/2 points (graded)

Random variables X, Y take on values in the range $\{-1, 0, 1\}$ and have the following joint distribution.

		$Y=-1$	$Y=0$	$Y=1$
X	-1	0	0	$1/3$
	0	0	$1/3$	0
	1	$1/3$	0	0

a) What is the covariance between X and Y?

-2/3

$-\frac{2}{3}$

Answer: -2/3

b) What is the correlation between X and Y?

-(2/3) / (sqrt(2/3) * sqrt(2

$-\frac{\frac{2}{3}}{\sqrt{\frac{2}{3}} \cdot \sqrt{\frac{2}{3}}}$

Answer: -1

? Hint (1 of 1): The first step is to compute the distribution of X on its own and of Y on its own.

Next Hint

Submit

Answers are displayed within the problem

Problem 18

2/2 points (graded)

Random variables X, Y take on values in the range $\{-1, 0, 1\}$ and have the following joint distribution.

		$Y=-1$	$Y=0$	$Y=1$
X	-1	$1/6$	0	$1/6$
	0	0	$1/3$	0
	1	$1/6$	0	$1/6$

a) Are X and Y independent?

- (•) dependent
- ( ) independent

✔

b) Are X and Y uncorrelated?

- ( ) correlated
- (•) uncorrelated

✔

Submit

Answers are displayed within the problem

Problems 19-20 correspond to "Two-dimensional generative modeling with the bivariate Gaussian"

Problem 19

2/2 points (graded)

Each of the following scenarios describes a joint distribution (x, y). In each case, give the parameters of the (unique) bivariate Gaussian that satisfies these properties.

a) x has mean 2 and standard deviation 1, y has mean 2 and standard deviation 0.5, and the correlation between x and y is -0.5.

- ○ $\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$
- ○ $\mu = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & \frac{1}{2} \end{pmatrix}$
- ● $\mu = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix}$

✔

b) x has mean 1 and standard deviation 1, and y is equal to x.

- ○ $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- ● $\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
- ○ $\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$
- ○ $\mu = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

✔

Submit

Answers are displayed within the problem

Problem 20

3/3 points (graded)

Here are four possible shapes of Gaussian distributions:

For each of the following Gaussians $N(\mu, \Sigma)$, indicate which of these shapes (1,2,3,4) is the best approximation.

a) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$

1

✔ **Answer:** 1

1

b) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 2 \\ 2 & 1 \end{pmatrix}$

3

✔ **Answer:** 3

3

c) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix}$

2

✔ **Answer:** 2

2

? **Hint (1 of 1):** These distributions all have mean zero, so only the covariance matrix matters. The thing to bear in mind is that the Gaussian is tilted up if and only if the covariance between the two features (the off-diagonal entry in the matrix) is positive, and it is tilted down if and only if the covariance is negative.

Next Hint

Submit

Answers are displayed within the problem