

Midterm Exam Guide: Solutions to Example Problems

1. (a) Probability of being male, given that you have autism spectrum disorder.

$$P(M|A) = 0.8 \text{ (given in problem).}$$

- (b) Probability of having autism spectrum disorder, given that you are a male. Using Bayes' Rule:

$$P(A|M) = \frac{P(M|A)P(A)}{P(M)} = \frac{0.8 \times 0.006}{0.5} = 0.0096$$

(c)

$$\begin{aligned} P(M^c \cap A) &= P(A - M) && \text{definition of set minus} \\ &= P(A) - P(A \cap M) && \text{difference rule} \\ &= P(A) - P(M|A)P(A) && \text{multiplication rule} \\ &= 0.006 - 0.8 \times 0.006 \\ &= 0.0012 \end{aligned}$$

2. We need total probability. Let R_1 be the event that you first select a red sock. Then R_1^c is the event that you first select a blue sock. We have

$$P(R_1) = \frac{2}{3}, \quad P(R_1^c) = \frac{1}{3}$$

Now, let R_2 be the event that the second sock you select is red. The conditional probabilities given the first sock that we need are:

$$P(R_2 | R_1) = \frac{3}{5}, \quad P(R_2^c | R_1^c) = \frac{1}{5}$$

Putting it together, we have

$$\begin{aligned} P(\text{"match"}) &= P(\text{"two reds"}) + P(\text{"two blues"}) \\ &= P(R_1 \cap R_2) + P(R_1^c \cap R_2^c) \\ &= P(R_2 | R_1)P(R_1) + P(R_2^c | R_1^c)P(R_1^c) \\ &= \frac{3}{5} \times \frac{2}{3} + \frac{1}{5} \times \frac{1}{3} \\ &= \frac{7}{15} \end{aligned}$$

(Note: you wouldn't have to do the last step simplifying the fraction on the exam. It's fine to just plug in the numbers that are in the second-to-last step and leave it there.)

3. (a) Naïve Bayes assumes that the variables for each dimension are independent. Or in other words, we can write the likelihood function as a product of univariate probability density functions:

$$p(x^1, x^2, \dots, x^d | c) = p(x^1 | c)p(x^2 | c) \cdots p(x^d | c)$$

- (b) This assumption is not true when there is a dependency between the x^i variables. Sometimes dependency (or correlation) may not be a problem. However, for naïve Bayes to work, there must be some separation between the classes in the single x^i variables. It will fail when the two classes are not well separated in any of the single x^i dimensions, but they are separated in the full-dimensional space.
 - (c) The training phase fits the model to the data, which means that we estimate the parameters of the normal distributions. There is a mean (μ) and variance (σ^2) parameter for each x^i given $c = \text{"dog"}$ and $c = \text{"cat"}$, giving $2d$ means and $2d$ variances to estimate. The μ parameters are the averages of the x^i in a given class, and the σ^2 parameters are the corresponding variances of that data.
4. (a) The null hypothesis is that the drug has no effect on making your recovery time shorter.
- (b) The correct answer is iv.
- (c) It looks effective! The drug has 3 to 1 chance of fast recovery, whereas the placebo has even chances of fast or slow recovery. Of course, to be sure, we'd have to do the analysis and get a p value or a Bayesian posterior probability.