-CSE 242 Homework 1-

Charles Dickens cadicken@ucsc.edu,
Kejun Chen kchen158@ucsc.edu
October 2019

Problem 1)

Hack the Titanic (2 pts). Kaggle hosts a number of machine learning competitions, one of the introductory ones is predicting survivorship of Titanic passengers. See the contest website at `https://www.kaggle.com/c/titanic/overview` (you may have to register to get full access). Look at the overview, data, evaluation metrics, rules, and public leaderboard. How good are the top teams on the leaderboard? Does this seem reasonable? Notice that the test data is split into two parts: one for the public leaderboard and one for the final evaluation. Assume that the same 50% of the test data is used to evaluate the submissions on the public leaderboard, and that you receive your leader- board score (performance on the public leaderboard half of the test data) with each submission. Describe how, with patience, you can get a 100% score on the leader- board without doing any real machine learning (description sufficient, no need to make Kaggle submissions).

- How good are the top teams on the leaderboard?
  The teams on the top of the public leaderboard are getting 100% accuracy on the 50% of the test data being used...!
  - Does this seem reasonable?
    100% accuracy is ***not*** reasonable, this is tells me there is some information being leaked from the test set that is being used to train the model.
- Describe how, with patience, you can get a 100% score on the leaderboard without doing any real machine learning.
  With patience, one might use the leaderboard score that is given at each submission to learn a mapping from each test observation id to the correct label.
  We are given the unlabeled test data, which is 418×11, and we know that 50% of the test data is used to calculate the public leaderboard score. Therefore, all we must do is try adjusting a random mapping of the features PassengerId to Survided for each example one at a time. For example, we can start with the first test observation and change the starting prediction by changing that single survival outcome and see if our leaderboard score improves, if no change, then that example is in the final set, else, we can determine the true outcome of that passenger. This will take at most 419 submissions, and our "model" will have 100% accuracy.

Problem 2)

My new neighbors have two children. Assume that the gender of the child is like a coin flip, so the four possibilities (oldest child first), GG, GB, BG, BB, are all equally likely (using G for girl, B for boy).

1. Suppose I ask them if they have any boys, and they say yes. What is the probability that one of their two children is a girl? (hint: assume that the information is accurate, what is the conditioning induced by this information?)

   The information we are given implies that at least one of the two children are boys, so, we have that

$$P(\{GG, GB, BG\} \mid \{GB, BG, BB\}) = \frac{P(\{GG, GB, BG\} \cap \{GB, BG, BB\})}{P(\{GB, BG, BB\})} = \frac{P(\{GB, BG\})}{P(\{GB, BG, BB\})}$$

$$= \frac{P(\{GB\}) + P(\{BG\})}{P(\{GB\}) + P(\{BG\}) + P(\{BB\}))} = \frac{1/4 + 1/4}{1/4 + 1/4 + 1/4} = \frac{2}{3}$$

2. Suppose instead that I happen to see one of the children run by, and it is a boy. What is the probability that the other child is a girl? (Hint: it may help to first consider the case where you saw the oldest child. Is the case where you see the youngest child the same? Does the answer depend on the probability that you see the oldest child?)

First consider the case that we know we saw the youngest child. Then we have

$$P(\text{"Other child is girl"} \mid \text{"youngest is boy"})$$

$$= P(\{GG, GB, BG\}|\{GB, BB\}) = \frac{P(\{GG, GB, BG\} \cap \{GB, BB\})}{P(\{GB, BB\})} = \frac{P(\{GB\})}{P(\{GB, BB\})}$$

$$= \frac{P(\{GB\})}{P(\{GB\}) + P(\{BB\}))} = \frac{1/4}{1/4 + 1/4} = \frac{1}{2}$$

Now consider the case that we know we saw the oldest child. Then we have

$$P(\text{"Other child is girl"} \mid \text{"oldest is boy"})$$

$$= P(\{GG, GB, BG\}|\{BG, BB\}) = \frac{P(\{GG, GB, BG\} \cap \{BG, BB\})}{P(\{BG, BB\})} = \frac{P(\{BG\})}{P(\{BG, BB\})}$$

$$= \frac{P(\{BG\})}{P(\{BG\}) + P(\{BB\}))} = \frac{1/4}{1/4 + 1/4} = \frac{1}{2}$$

Then, we have, that

$$P(\text{"Other child is girl"})$$

$$= P(\text{"Other child is girl"} \mid \text{"youngest is boy"})P(\text{"youngest is boy"}) +$$

$$P(\text{"Other child is girl"} \mid \text{"oldest is boy"})P(\text{"oldest is boy"})$$

$$= 1/2P(\text{"youngest is boy"}) + 1/2P(\text{"oldest is boy"})$$

$$= 1/2(P(\text{"youngest is boy"}) + P(\text{"oldest is boy"})) = 1/2(1) = 1/2$$

Therefore, $P(\text{"Other child is girl"}) = 1/2$ regardless of the prior on whether you saw the youngest or oldest child.

Problem 3)

Consider a prediction problem where we are supposed to predict if it is safe to enter an intersection or not, based on the features available to an autonomous vehicle. Assume that the features are rich enough so that some formula over the features is always correct. In other words, for the safe combinations of features you always want to enter the intersection, and for the unsafe combinations of features you never want to enter the intersection. Assume also that there is some unknown distribution D over the feature vectors generated when approaching an intersection (we model each approach to an intersection as generating a feature vector drawn independently from D) and we have a rule predicting safe or unsafe that is 90% correct (on randomly drawn feature vectors).

1. What is the probability that, on a randomly drawn test set of 10 examples, the rule appears perfect?

   Given each approach to an intersection is equivalent to generating a feature vector drawn independently from D and we have a rule predicting safe or unsafe that is 90% accurate, we can model rule predictions on a randomly drawn test set as Bernoulli trails with a probability of success $p = 0.9$. Thus on a randomly drawn test set of 10 examples, the rule appears perfect with probability

   $$p^{10} = 0.9^{10} \approx 0.35$$

2. What is the probability that, on a randomly drawn test set of 50 examples, the rule appears perfect?

   With the same reasoning from the previous part of this problem we find that on a randomly drawn test set of 50 examples, the rule appears perfect with probability

   $$p^{50} = 0.9^{50} \approx 0.0052$$

3. Assume a class of 100 students each independently draws 50 examples to test the rule, what is the probability that at least one of the students draws a test set that makes the rule appear perfect?

   We can model perfect rule accuracies on a students test set as Bernoulli trails with a probability of success $p = 0.0052$. Thus, out of 100 students, the probability that at least one student draws a test set that makes the rule appear perfect.

$$p(\text{"at least one perfect"}) = 1 - p(\text{"no Perfect"}) \approx 1 - (1 - 0.0052)^{100} \approx 0.41$$

Problem 4)

Two random variables V and W are independent if for all values v and w, the events $(V = v)$ and $(W = w)$ are independent, so the joint density (or distribution) $p(V = v, W = w) = p(V = v)p(W = w)$.

1. Prove that if V and W are independent random variables then the expectation of their product, $E[VW]$, equals the product of their expectations, $E[V]E[W]$. Assume that the random variables V and W are integer valued, so that the distributions $p(V)$ and $p(W)$ are distributions over the integers and the expectations can be written as sums.

   For two discrete random variables:

$$\begin{aligned}
E[VW] &= \sum_{i,j} \left(v_i w_j p(V = v_i, W = w_j)\right) \\
&= \sum_{i,j} \left(v_i w_j p(V = v_i) p(W = w_j)\right) \\
&= \sum_{i} \sum_{j} \left(v_i w_j p(V = v_i) p(W = w_j)\right) \\
&= \sum_{i} \left(v_i p(V = v_i)\right) \sum_{j} \left(w_j p(W = w_j)\right) \\
&= E(V)E(W)
\end{aligned}$$

   For two continuous random variables:

$$\begin{aligned}
E[VW] &= \int_i \int_j vwp(vw) \\
&= \int_i \int_j vwp(v)p(w) \\
&= \int_i vp(v) \int_j wp(w) \\
&= E(V)E(W)
\end{aligned}$$

2. Find an example of two random variables S and T (i.e.describe their joint distribution) that are not independent and verify that that $E[T]E[S] \neq E[TS]$.

   For two random variables S and T, assume they are linear dependent. The random variable T is totally determined by S and satisfy $T = S$. Assume random variable S satisfies discrete uniform distribution and can have two integer values, $S = \{0, 1\}$. The random variables distribution of T and S:

S	0	1
$Probability(S)$	1/2	1/2

T	0	1
$Probability(T)$	1/2	1/2

   Through their distributions, we can easily know $E(S) = \frac{1}{2}$, $E(T) = \frac{1}{2}$.

   We can also get their joint distribution probability:

$T \backslash S$	0	1
0	1/2	0
1	0	1/2

Similar, we can obtain $E(ST) = 1 * 1 * \frac{1}{2} = \frac{1}{2}$. Obviously, $E(S)E(T) = \frac{1}{4}$, so $E(ST) \neq E(S)E(T)$

3. Describe a joint distribution for two (integer-valued) random variables Q and R where where Q and R are not independent, but where $E[Q]E[R] = E[QR]$ (and justify that the Q and R have these properties).

   Assume random variable Q satisfies discrete uniform distribution and can have three integer values, $Q = \{-1, 0, 1\}$. The random variable R satisfies $R = Q^2$, then R is completely determined by Q. Therefore, Q and R are not independent. The random variables distribution of Q and R:

Q	-1	0	1
$Probability(Q)$	1/3	1/3	1/3

T	0	1
$Probability(R)$	1/3	2/3

Through their distributions, we can easily know $E(Q) = 0$, $E(R) = \frac{2}{3}$.

We can also get their joint distribution probability:

$R \backslash Q$	-1	0	1
0	0	1/3	0
1	1/3	0	1/3

Similar, we can obtain $E(QR) = -1 * 1 * \frac{1}{3} + 1 * 1 * \frac{1}{3} = 0$. Obviously, $E(QR) = E(Q)E(R) = 0$, but the two random variables are not independent.