**P1:** Using a little bit of algebra, prove that (4.2) is equivalent to (4.3). In other words, the logistic function representation and logit representation for the logistic regression model are equivalent.

**Ans:**

|  |  |
| --- | --- |
|  | (4.2) |
|  | (4.3) |

Accordingly,

Now,

|  |  |
| --- | --- |
|  | **[proved]** |

**P4:** When the number of features is large, there tends to be a deterioration in the performance of KNN and other *local* approaches that perform prediction using only observations that are *near* the test observation for which a prediction must be made. This phenomenon is known as the *curse of dimensionality*, and it ties into the fact that non-parametric approaches often perform poorly when is large. We will now investigate this curse.

1. Suppose that we have a set of observations, each with measurements on feature, . We assume that is uniformly (evenly) distributed on [0, 1]. Associated with each observation is a response value. Suppose that we wish to predict a test observation’s response using only observations that are within 10% of the range of closest to that test observation. For instance, in order to predict the response for a test observation with = 0.6, we will use observations in the range [0.55, 0.65]. On average, what fraction of the available observations will we use to make the prediction?

**Ans:** If , then the observations lie in the interval wherein the interval length is .

Now, if , then the observations lie in the interval  which represents a fraction of . Similarly, if , then the observations lie in the interval  or in the ( fraction. Therefore, the average fraction to be used for prediction can be calculated by solving the following integration:

**So, the fraction of observations available for prediction is 9.75%.**

1. Now suppose that we have a set of observations, each with measurements on features, and . We assume that (, ) are uniformly distributed on [0, 1] [0, 1]. We wish to predict a test observation’s response using only observations that are within 10% of the range of and within 10% of the range of closest to that test observation. For instance, in order to predict the response for a test observation with = 0.6 and = 0.35, we will use observations in the range [0.55, 0.65] for and in the range [0.3, 0.4] for . On average, what fraction of the available observations will we use to make the prediction?

**Ans:** Here, we assume and to be mutually independent and then use the calculation from P4 (a). Therefore, the fraction of available observations will be (9.75% 9.75%) **0.95%** (rounded to two decimal places).

1. Now suppose that we have a set of observations on features. Again, the observations are uniformly distributed on each feature, and again each feature ranges in value from 0 to 1. We wish to predict a test observation’s response using observations within the 10% of each feature’s range that is closest to that test observation. What fraction of the available observations will we use to make the prediction?

**Ans:** Assuming all the features are mutually independent like in P4 (b), the fraction of available observations is . Therefore, there are no observations available for prediction.

1. Using your answers to parts (a)–(c), argue that a drawback of KNN when is large is that there are very few training observations “near” any given test observation.

**Ans:** As calculated in P4 (a)-(c) we see that, if , then, . Therefore, when is large, there are very few training observations “near” any given test observation. This is a drawback of the KNN approach.

1. Now suppose that we wish to make a prediction for a test observation by creating a -dimensional hypercube centered around the test observation that contains, on average, 10% of the training observations. For = 1, 2, and 100, what is the length of each side of the hypercube? Comment on your answer.

**Ans:** Let, the length of each side of the hypercube be . Now, following the calculation done for the 10% of the training observations in P4 (a), when . And, in general, for where is a positive integer representing the dimension of the hypercube. In the following table, the values (rounded to two decimal places) for are shown for = 1, 2, and 100.

|  |  |
| --- | --- |
|  |  |
| 1 | 0.1 |
| 2 | 0.32 |
| 100 | 0.98 |

So, the length of each side of the hypercube increases with the increase in its dimensionality. This means that as increases, for 10% of the training observations, we need to include the entire range of each feature.

**P6:** Suppose we collect data for a group of students in a statistics class with variables = hours studied, = undergrad GPA, and = receive an A. We fit a logistic regression and produce estimated coefficient, = −6, = 0.05, = 1.

1. Estimate the probability that a student who studies for 40 h and has an undergrad GPA of 3.5 gets an A in the class.

**Ans:** Accordingly, for logistic regression,

where, is the estimated probability, and .

Here,

(rounded to two decimal places).

Therefore, the estimated probability is **0.38**.

1. How many hours would the student in part (a) need to study to have a 50% chance of getting an A in the class?

**Ans:** Accordingly,

where, we have to solve for = hours studied.

We know from P1 that,

Therefore, the student has to study for **50 h**.

**P9:** This problem has to do with odds.

1. On average, what fraction of people with an odds of 0.37 of defaulting on their credit card payment will in fact default?

**Ans:** Accordingly,

where is the probability of defaulting on credit card payment.

Now,

So, on average, **27%** of the people default on their credit card payment.

1. Suppose that an individual has a 16% chance of defaulting on her credit card payment. What are the odds that she will default?

**Ans:** Similar to P9 (a),

So, the odds that she will default is **0.19**.