

Assignment 54-2

riley.r.paddock

October 2020

Setup

Suppose you want to estimate the probability that you will get into a particular competitive college. You had a bunch of friends a year ahead of you that applied to the college, and these are their results:

- Martha was accepted. She was the 95th percentile of her class, got a 33 on the ACT, and had an internship at a well-known company the summer before she applied to college.
- Jeremy was rejected. He was in the 95th percentile of his class and got a 34 on the ACT.
- Alphie was accepted. He was in the 92nd percentile of his class, got a 35 on the ACT, and had agreed to play on the college's basketball team if accepted.
- Dennis was rejected. He was in the 85th percentile of his class, got a 30 on the ACT, and had committed to run on the college's track team if accepted.
- Jennifer was accepted. She was in the 80th percentile of her class, got a 36 on the ACT, and had a side business in 3D printing that was making \$ 15,000 per year.
- Martin was rejected. He was in the 85th percentile of his class, got a 29 on the ACT, and had was a finalist in an international science fair.
- Mary was accepted. She was in the 95th percentile of her class, got a 36 on the ACT, and was a national finalist in the math olympiad.
- Dean was rejected. He was in the 87th percentile of his class, got a 31 on the ACT, and was a national finalist in the chemistry olympiad.

- Adam was accepted. He was in the 99th percentile of his class and got a 36 on the ACT.
- Jeremy was rejected. He was in the 95th percentile of his class and got a 32 on the ACT.

Part A.

Create a quantitative dataset to represent this information, and include it in your writeup. Name your features appropriately.

```
['Name', 'Accept/Reject', 'Percentile', 'ACT', 'Extracurricular']
```

```
['Martha', 'Accepted', 95, 33, 'Internship']
['Jeremy', 'Rejected', 95, 34, 'None']
['Alphie', 'Accepted', 92, 35, 'Basketball']
['Dennis', 'Rejected', 85, 30, 'Track']
['Jennifer', 'Accepted', 80, 36, 'Personal buisness']
['Martin', 'Rejected', 85, 29, 'National Science Fair']
['Mary', 'Accepted', 95, 36, 'Math Olympiad']
['Dean', 'Rejected', 87, 31, 'Chemistry Olympiad']
['Adam', 'Accepted', 99, 36, 'None']
['Jeremy', 'Rejected', 95, 32, 'None']
```

Part B.

Decide what type of model you will use to model the probability of acceptance as a function of the features in your dataset. State and justify the form of the model in your writeup.

A logistic model will work well because it will give us a number between 0 and 1 so we can know our odds of getting into the college. We will create 1 term for percentile, ACT, and Extracurricular. We wil also have interaction terms because high percentile and low ACT may mean they just went to a small school and maybe a low ACT is made up for by an extra curricular. We will treat the extracurricular as a general addition because any student who had an extracurricular and got rejected had an ACT of below 30 or below and a percentile below 90. So the specificity of the extracurricular is not as important as the general idea that they had one.

Part C.

Fit the model to the data. For each feature, answer the following questions:

According to your model, as that variable increases, does the estimated probability of acceptance increase or decrease? Does that result make sense? If so, why? (If not, then something is wrong with your model, and you need to figure out what's going wrong.)

Term 1: Percentile

The estimated probability increases the higher the percentile is. This makes sense because the higher the percentile the better of a student you are.

Term 2: ACT

The estimated probability increases the higher the ACT score is. This makes sense because the ACT is a representation of your intelligence and the higher you are the more likely you are to succeed at college.

Term 3: Extracurricular

The estimated probability increases if you have an extra curricular. This makes sense because if you do an extracurricular you show initiative that colleges like

Term 4: Percentile*ACT

The estimated probability decreases as this variable increases. This makes sense because sometimes Percentile can be deceiving based on the size of the school. So by adding an interaction between a standard variable, like ACT, and the Percentile it relieves some of the deceiving properties of Percentile.

Term 5: Percentile*Extracurricular

The estimated probability decreases as this variable increases. This makes sense because if a student has a high percentile and an extracurricular may seem off and may be another sign of the deceiving properties of percentile because it is more difficult to keep up good grades and an extra curricular. However if a student is doing this it is accounted for in the individual extracurricular and percentile variables. This is just a small buffer to help alleviate some of the deception of Percentile

Term 6: ACT*Extracurricular

The estimated probability increases as the ACT*Extracurricular increases. This makes sense because if you have an extracurricular they consider you more and thus give your ACT score a little more weight.

Part D.

Estimate the probability of being accepted for each of the data points that you used to fit the model. How well does this match up with reality?

Student(Acceptance): Percent Chance of getting in from Model

Martha(Accepted):0.9966895600782484
Jeremy(Rejected): 0.0009999946440241599
Alphie(Accepted): 0.9979681707761708
Dennis(Rejected): 0.000986149161169717
Jennifer(Accepted): 0.9983507182902432
Martin(Rejected): 0.00010058029388191834
Mary(Accepted): 0.9994654728267108
Dean(Rejected): 0.056700032889175724
Adam(Accepted): 0.9989999940672725
Jeremy(Rejected): 0.0009999949836823289

Part E.

Estimate your probability of being accepted if you are in the 95th percentile of your class and got a 34 on the ACT. Justify why your model's prediction is reasonable.

A student in the 95th percentile with an ACT of 34 and no extracurricular has a probability of 0.0009999946440241599 of getting into college. This is reasonable because I had no extracurricular which have a large influence on the acceptance at this particular college.

Part F.

Now suppose that you have an opportunity to do an internship at a well-known company the summer before you apply to college. If you do it, what will your estimated probability of acceptance become? Based on this information, how much does the internship matter in terms of getting into the college you want? If I had an internship prior to applying my probability of getting in will increase to 0.9981962881896385. So almost guaranteed. So obviously having an extra curricular, or an internship rather, is a huge in terms of getting into the college I want.