The influence of teaching styles on the scores of a diverse group of student

Jerry Chen, Aniruddha Prasad

Motivation

- We are both students from very different ethnic and socioeconomic backgrounds
- Both come from different two very different places each with their own style of teaching
- Students deserve quality education regardless of who they are or where they are from

Our Objectives

- Analyze the 8th grade test score dataset and recommend a teaching method (standards-based or traditional)
- Find the best model for predicting scores and teachers
  - Scores: use teacher, socioeconomic, gender, and ethnic background
  - Teacher: use scores, socioeconomic, gender, and ethnic background
- Tune the model for higher accuracy
- Develop a “master” model with the highest predictibility
- Finally see if our hypotheses were true

The Dataset

"Student's Math Score for Different Teaching Styles"

- 215 observations
- Factors:
  - Teachers
    - Ruger: Standard-based learning
    - Wesson: Traditional learning
  - Gender
  - Ethnicity
  - Free reduced lunch
  - Score
- Structure type:
  - Float64: student, score
  - Object: teacher, gender, ethnic, freeredu
- Target variables:
  - Score & Teacher

	Student	Teacher	Gender	Ethnic	Freeredu	Score
0	1.0	Ruger	Female	Asian	Free lunch	76.0
1	2.0	Ruger	Female	Hispanic	Paid lunch	56.0
2	3.0	Ruger	Female	African-American	Free lunch	34.0
3	4.0	Ruger	Female	Asian	Paid lunch	59.0
4	5.0	Ruger	Male	Hispanic	Free lunch	73.0
...	...	...	...	...	...	...
211	212.0	Wesson	Male	African-American	Paid lunch	56.0
212	213.0	Wesson	Male	Hispanic	Free lunch	94.0
213	214.0	Wesson	Male	Hispanic	Paid lunch	91.0
214	215.0	Wesson	Female	African-American	Paid lunch	53.0
215	216.0	Wesson	Male	Hispanic	Paid lunch	57.0

```
Student      float64
Teacher       object
Gender        object
Ethnic        object
Freeredu      object
Score        float64
dtype: object
```

Some Hypotheses

- Expect higher performance with the standard-based method
- Expect higher performance with better socioeconomic status.
- Expect lower performance under Wesson with students who are underprivileged and not Caucasian
- Expect lower performance from female students with a worse socioeconomic status compared to male students.

Methods

Preprocessed raw data into individual classes for easier analysis. One-hot encoding was not used.

Regression Model accuracy
- KNN - 27.20%
- Multivariate Regression - 58.36%
- ANN

Classification Models accuracy - Some comments on this
- KNN
- SVC
- Random Forest Classifier

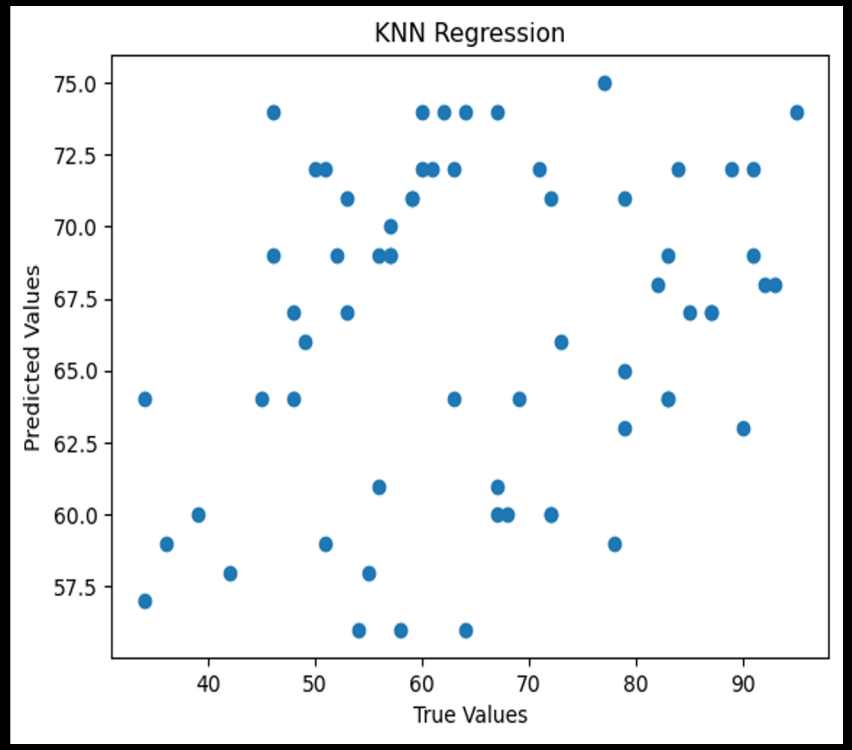
	Student	Teacher	Gender	Ethnic	Freeredu	Score
0	1.0	Ruger	Female	Asian	Free lunch	76.0
1	2.0	Ruger	Female	Hispanic	Paid lunch	56.0
2	3.0	Ruger	Female	African-American	Free lunch	34.0
3	4.0	Ruger	Female	Asian	Paid lunch	59.0
4	5.0	Ruger	Male	Hispanic	Free lunch	73.0
...	...	...	...	...	...	...
211	212.0	Wesson	Male	African-American	Paid lunch	56.0
212	213.0	Wesson	Male	Hispanic	Free lunch	94.0
213	214.0	Wesson	Male	Hispanic	Paid lunch	91.0
214	215.0	Wesson	Female	African-American	Paid lunch	53.0
215	216.0	Wesson	Male	Hispanic	Paid lunch	57.0

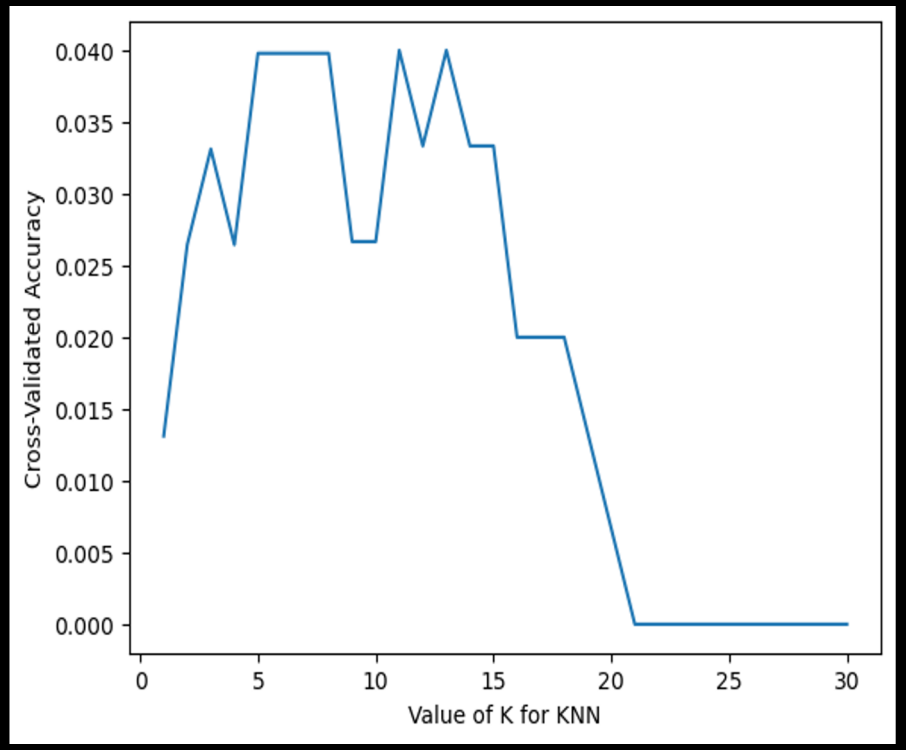
Results & Discussion

- Correlation matrix shows little correlation among variables -> low accuracy from our various models.
- Closer to 1 = higher correlation.
- Score and Teacher variables show the highest correlation, but still low.

	Student	Teacher	Gender	Ethnic	Freeredu	Score
Student	1.00	0.46	-0.03	0.03	0.00	0.37
Teacher	0.46	1.00	0.04	0.07	0.08	0.36
Gender	-0.03	0.04	1.00	-0.06	0.01	-0.13
Ethnic	0.03	0.07	-0.06	1.00	0.03	-0.01
Freeredu	0.00	0.08	0.01	0.03	1.00	-0.04
Score	0.37	0.36	-0.13	-0.01	-0.04	1.00

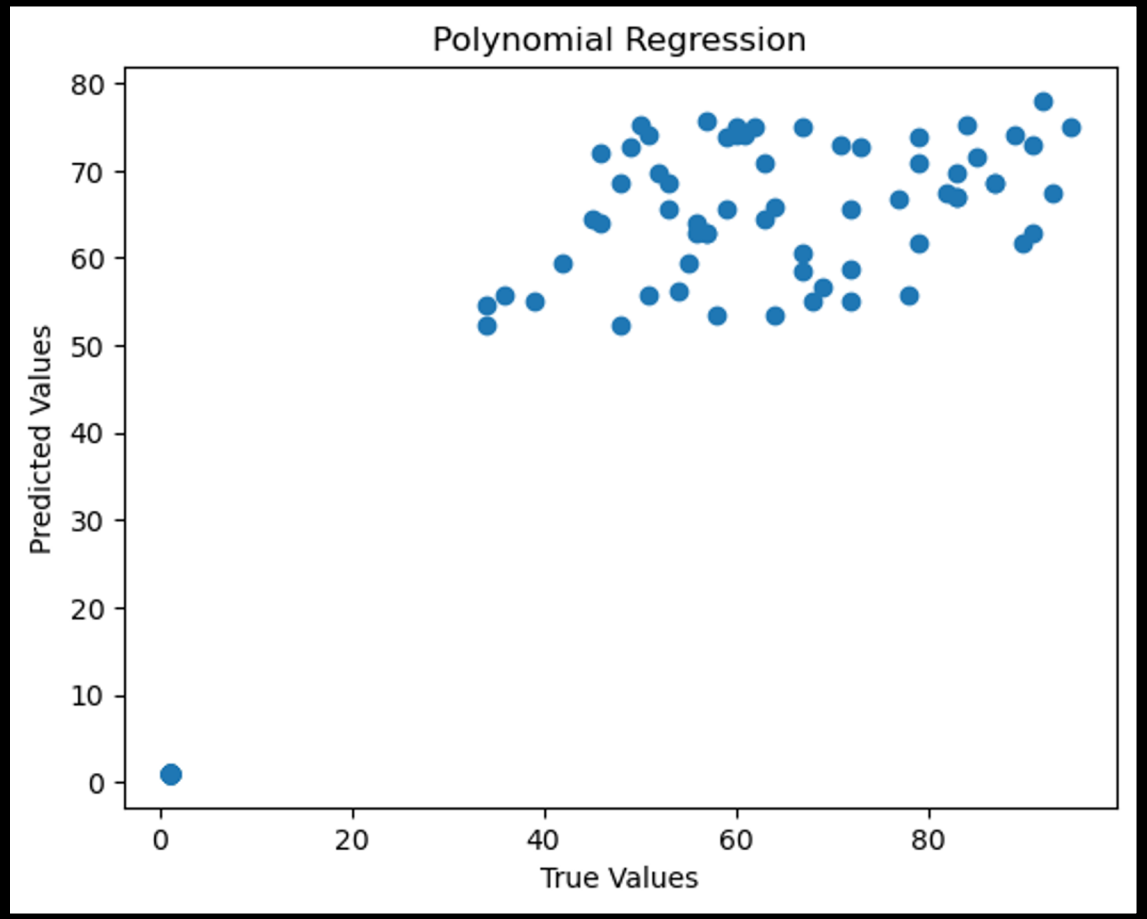
Failure of KNN on this Dataset





- Knn regression had abysmal performance with poor fit and poor prediction capabilities. (image on the left)
- A grid search (image on the right) revealed absolutely low scores for the KNN regressor hence why it was scrapped in favor of polynomial regression.

Performance of polynomial regressor



- Polynomial regressor had the highest accuracy on this dataset that had low correlation.
- Figure on the left shows this prediction capability. Displays higher accuracy and better prediction on the test set.

Results & Discussion

- Multivariate Regression had the highest accuracy -> chosen to predict score.
- Wesson has a higher average. However, we are still improving the accuracy of our models.
- No correlation between gender, ethnicity, and socioeconomics.

Predicted Score	Teacher	Gender	Ethnicity	Economic
63.89	Wesson	Male	African American	Free lunch
70.97	Wesson	Male	African American	Paid lunch
68.14	Wesson	Female	African American	Free lunch
65.59	Wesson	Female	African American	Paid lunch
62.75	Wesson	Male	Hispanic	Free lunch
69.83	Wesson	Male	Hispanic	Paid lunch
66.99	Wesson	Female	Hispanic	Free lunch
64.45	Wesson	Female	Hispanic	Paid lunch
61.61	Wesson	Male	Caucasian	Free lunch
68.69	Wesson	Male	Caucasian	Paid lunch
65.85	Wesson	Female	Caucasian	Free lunch
60.47	Wesson	Male	Asian	Free lunch
67.55	Wesson	Male	Asian	Paid lunch
64.71	Wesson	Female	Asian	Free lunch

Predicted Score	Teacher	Gender	Ethnicity	Economic
55.73	Ruger	Male	African American	Free lunch
62.81	Ruger	Male	African American	Paid lunch
59.97	Ruger	Female	African American	Free lunch
57.42	Ruger	Female	African American	Paid lunch
54.59	Ruger	Male	Hispanic	Free lunch
61.67	Ruger	Male	Hispanic	Paid lunch
58.83	Ruger	Female	Hispanic	Free lunch
56.28	Ruger	Female	Hispanic	Paid lunch
53.45	Ruger	Male	Caucasian	Free lunch
60.53	Ruger	Male	Caucasian	Paid lunch
57.69	Ruger	Female	Caucasian	Free lunch
52.31	Ruger	Male	Asian	Free lunch
59.38	Ruger	Male	Asian	Paid lunch
56.55	Ruger	Female	Asian	Free lunch