ClickToPrice: Incorporating Visual Features of Product Images in Price Prediction

Abhinav Maurya
Carnegie Mellon University
Pittsburgh, PA – 15213
ahmaurya@cmu.edu

Abstract

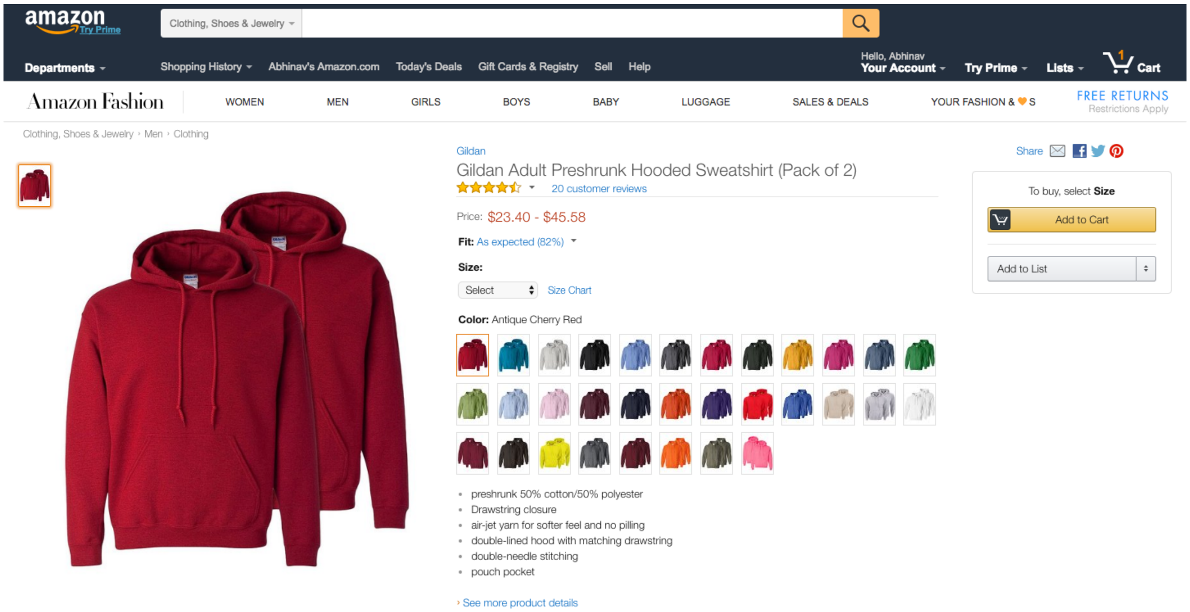
E-commerce websites provide product images in addition to textual description and other metadata about a product to shoppers. While the use of images in online product positioning is ubiquitous, its effect on shoppers and product prices has not been studied meaningfully in literature. In this paper, we make two primary contributions: (i) we find evidence of a causal link between the features of a product's image and its price, (ii) we conduct a comprehensive evaluation of regression algorithms to predict product price using metadata and image features of the product and find that the Random Forest algorithm provides the best prediction performance on this task. We discuss implications of our findings for e-commerce portals and mobile app creators, and present directions for future research to further investigate the predictive power and statistical significance of product images on their prices.

Keywords: Convolutional Neural Networks, Price Prediction

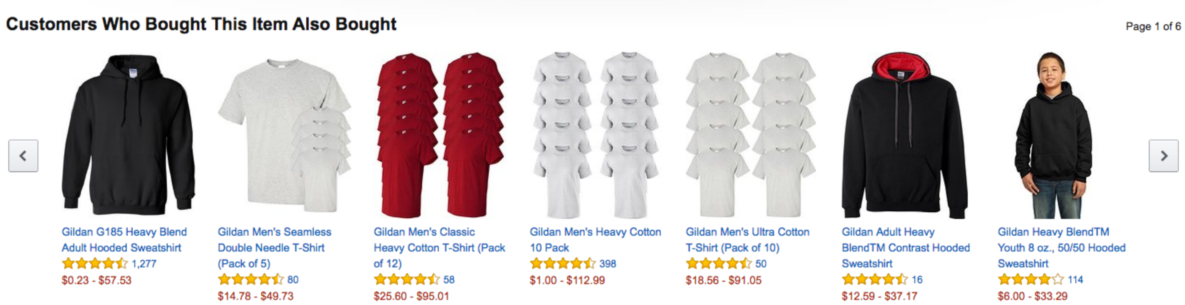
1. Introduction

E-commerce portals such as Amazon, Alibaba, and Flipkart widely use product images to position their products for sale. Product images are arguably the most important aspect of a product's page on an online shopping website. Their use in shaping consumer preferences and shopping decisions is ubiquitous. Figure (1) shows a few screenshots taken from an example product page on Amazon. It is difficult to miss how central product images are to the online shopping experience. However, little attention has been paid to understanding how images affect shopping behaviour and product prices, and how this can be used constructively by shopping portals to better market their products.

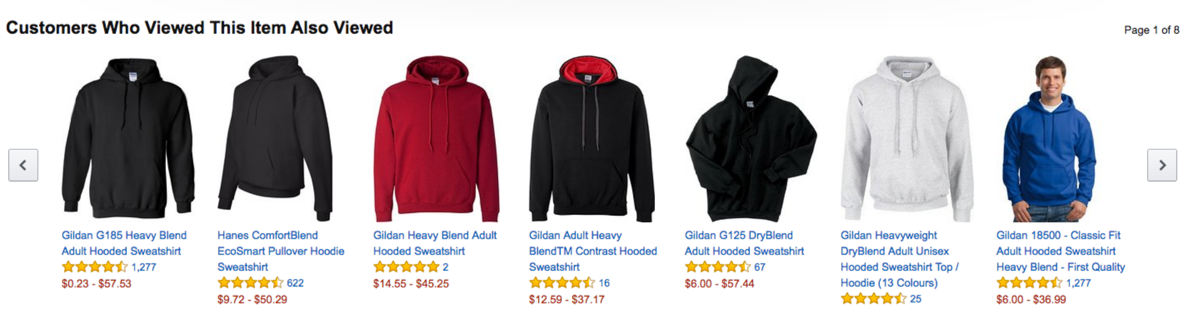
Our primary contribution is in investigating the relationship between visual features extracted from a product's image and the product's price. In this paper, we answer the following questions:-



(a) An example product page on Amazon



(b) Amazon website feature showing products bought together with current product



(c) Amazon website feature showing products viewed together with current product

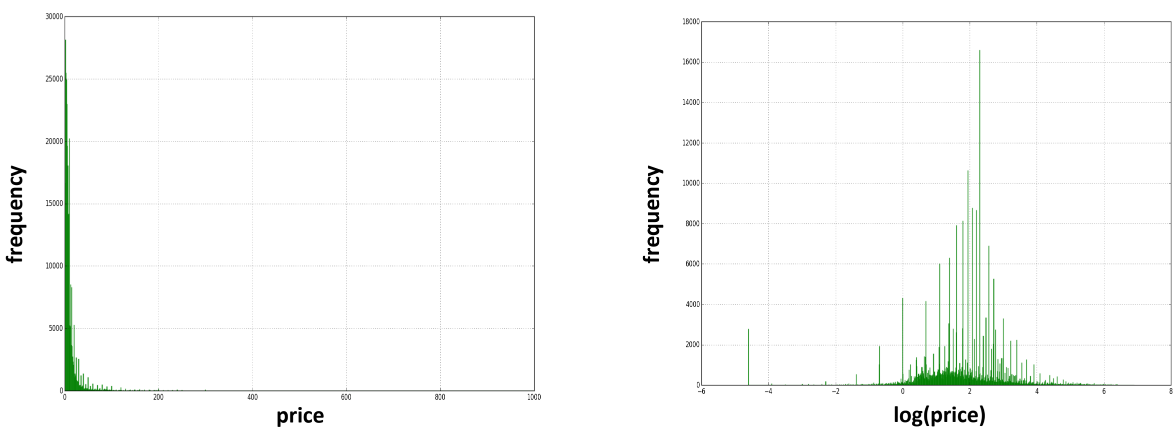
Figure 1: *Screenshots from the product page of a product titled "Gildan Adult Preshrunk Hooded Sweatshirt (Pack of 2)." The page is rich in product images, which affect a consumer's perception of the quality of the product.*

Table 1: *OLS Coefficents for Product Metadata*

Feature f_i	OLS coefficient β_i	p-value
SalesRank	-1.1879e-06	7.5344e-06
Average Price of Items Also Bought	0.3833	0.0
Average Price of Items Also Viewed	0.4631	0.0
Average Price of Items Bought Together	-0.2568	0.0
Average Price of Items Bought After Viewing	0.4874	0.0

Table 2: *Algorithms and Hyperparameters Evaluated*

Algorithm	Hyperparameters
Decision Tree Regression	Maximum Depth, Maximum Features (sqrt, log2, None)
Random Forest Regression	Number of Estimators, Maximum Features (sqrt, log2, None)
Gradient Boosting Regression	Number of Estimators, Maximum Features (sqrt, log2, None)
KNN Regression	Number of Neighbors
Ridge Regression	Regularization Parameter α



(a) Price Histogram

(b) log(Price) Histogram

Figure 2: *Descriptive statistics of target variable*

- **Predictive Power:** Do product images contain visual signals that can help us better predict product prices if they are not known? A potential application of such a predictive system can be in mobile apps for brick-and-mortar store shoppers. Such an app can allow the user to click product images and provide some metadata such as the type of product (e.g sunglasses). Using this information, the app can then provide a "guesstimate" of the product's price to help the shopper find the best deal on the product.

- **Causal Impact:** Another question related to product pricing using visual features is to determine if the connection between visual features and product prices is causal i.e. the relationship between visual features and

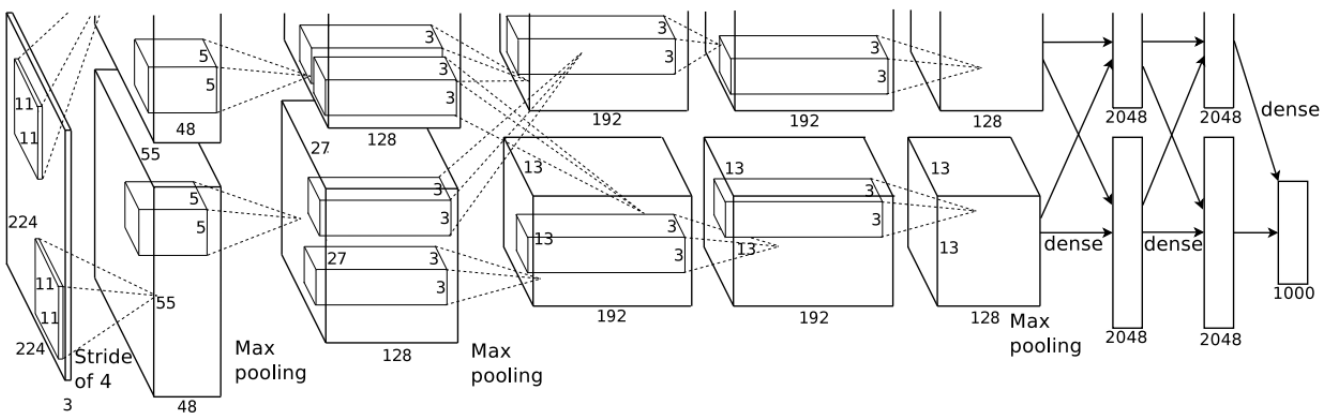


Figure 3: *CNN Architecture from [1]*

product price is statistically significant controlling for other information such as the product's category, its sales rank within its category, the sales prices of related products, etc. Such a causal link is essential to understand if the visual features are driving the pricing decisions of sellers and buyers in online marketplaces or are they just correlated with product prices through other observed features. If a causal link cannot be determined, then it would not be possible to plan constructive interventions such as the possibility of shoppers enhancing the product images to command a better price for the product in the online marketplace.

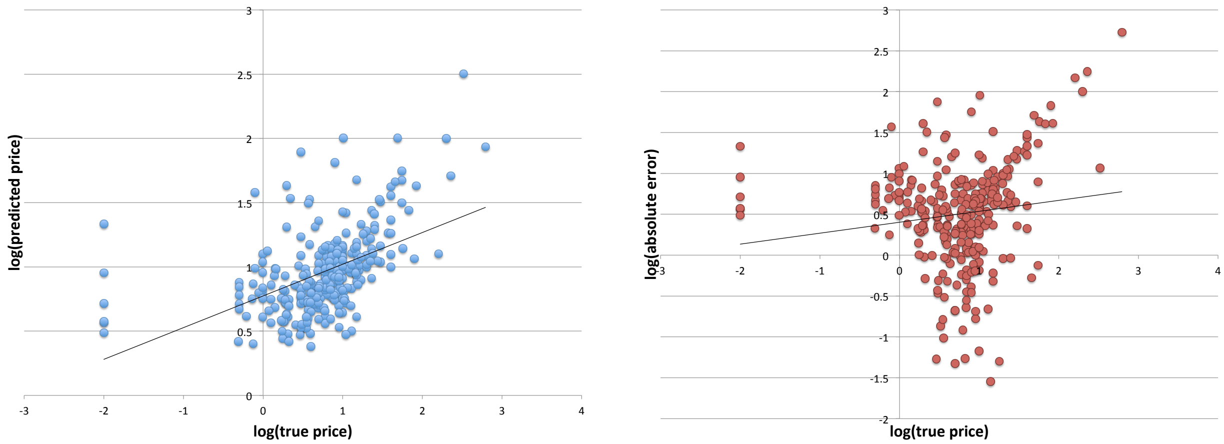
It is well-known that predictive performance of extracted features does not necessarily imply that the relationship between the features and the target variable is statistically significant. Hence, it is necessary to address both questions to exploit the relationship between visual features and product prices constructively.

2. Dataset

In order to answer our research question, we employ a freely available Amazon products dataset[1] used in [2] and [3]. The dataset was made available by the authors of these papers upon request. It consists of the following metadata for each product:-

- ASIN: An ID unique to each product within the Amazon online catalogue
- Title: A descriptive title that is used on the product's Amazon page as a prominent headline
- Categories: the list of categories a product belongs to e.g. Cellphones and Accessories

[1] http://jmcauley.ucsd.edu/data/amazon/



(a) True Price versus Predicted Price (log-log scale) (b) True Price versus Absolute Error (log-log scale)

Figure 4: *Analysis of the prediction results obtained from the best predictor, Random Forest with 200 maximum estimators.*

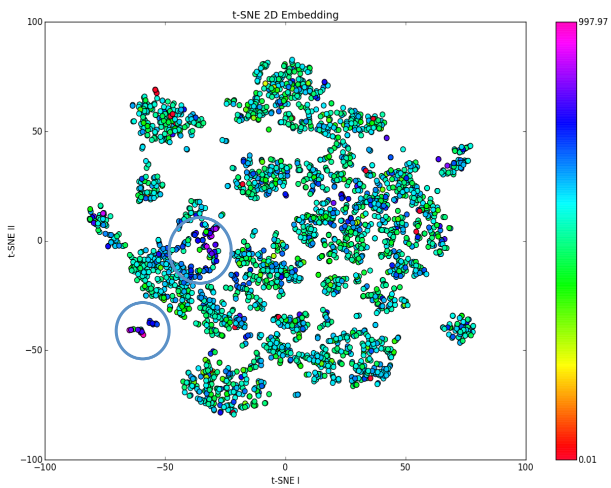


Figure 5: *t-SNE Visualization. Two blue circles indicate clusters of highly priced products.*

- Brand: name of the brand or manufacturer
- SalesRank: rank of the product within its category
- Products also bought: ASINs of products frequently bought by people who bought this product
- Products also viewed: ASINs of products frequently viewed by people who view this product
- Products bought together: ASINs of products frequently bought alongwith the product
- Products bought after viewing: ASINs of products bought after viewing this product

- Price: Price of the product as listed on Amazon at the time of crawling the product page

- Image URL: A URL for the product's image.

Reviews and ratings for all the products are also available as a separate dataset, though we have not used that information in this paper.

In addition to structured metadata and unstructured text, the dataset also provides visual features extracted from the last fully connected layer of a convolutional neural network (CNN) pretrained on 1.2 million ImageNet (ILSVRC2010) images. The structure of the CNN is shown in figure (3) which appears in [1]. The pretrained model is available as a reference model in the Caffe deep learning library [4]. Each product image yields 4096 CNN features from the FC7 layer of the ImageNet Caffe model. We consider that the predictive power of a product's images can be summarized using the visual features extracted from the pretrained ImageNet Caffe model.

True prices in the considered category "Cellphones and Accessories" range from $0.01 to $997.97. In figure (2.a), we show the histogram of true prices using 1000 bins. We can see that the distribution over prices follows a power-law. In figure (2.b), we plot the log of true prices which shows a clearer distribution over the bins.

3. Results

In this section, we demonstrate both the predictive power and causal impact of visual features in predicting prices of Amazon products. For this preliminary pilot investigation, we limited ourselves to the Amazon product category "Cellphones and Accessories," due to the size of the dataset and the need to perform a comprehensive evaluation of the regression algorithms.

3.1. Causal Impact of Visual Features

Before we use the visual features for prediction, we need to understand if the visual features add value to the task of product price prediction, controlling for all other information i.e. metadata about the product. In statistics and econometrics, causal inference is usually carried out by regressing the target variable against the independent variable of interest as well as other information we want to control to avoid detecting spurious correlational relationships. In our case, price of a product is the target variable, visual features of the product image are independent variables whose effect on price is to be tested, and the product-specific control variables include the following metadata:-

- Category: We have limited ourselves to just one category, and therefore controlled for it. Further categories will be analyzed independent of each other in future work.

- SalesRank
- Average price of products also bought
- Average price of products also viewed
- Average price of products bought after viewing
- Average price of products bought together

After constructing such a machine learning regression dataset, we run ordinary least squares (OLS) on it. The goal of running OLS is not to get good MSE. Instead, the focus of causal inference using OLS is on the statistical significance of the coefficients obtained using the OLS procedure. If the coefficient of an independent variable is statistically significant as per a two-sided t-test, the variable is considered to have causal impact on the target variable.

Our OLS run indicated that all metadata features such as SalesRank, and average prices of related products are highly statistically significant, as indicated in the table (1).

In addition to the statistical significance of product metadata features, we also found that **52 OLS coefficients corresponding to visual features were significant at 0.001 level, and 9 OLS coefficients corresponding to visual features were significant at 0.0001 level**. Thus, we conclude that even after controlling for product metadata, visual features are statistically significant in predicting prices. This implies that if the same product were advertised differently using better product images, it might fetch the seller a better price.

3.2. Predictive Power of Visual Features

We use the dataset described above to perform an evaluation of various regression algorithms on the task of price prediction. Each algorithm has certain hyperparameters which we also vary to get the best validation error on a heldout 10% validation dataset. We choose five regression algorithms to evaluate for the task as shown in table (2). We used the `scikit-learn` Python package for Machine Learning to evaluate the performance of these algorithms against each other. We found that Random Forest with no constraint on the maximum number of variables and trained by allowing for at least 200 base estimators gives the best performance. The best mean absolute error (MAE) obtained using the Random Forest regressor is around \$7.36. The baseline regressor which just outputs the mean price of all products in the category “Cellphones and Accessories” provides an MAE of \$12.65. Thus, Random Forest improves on the MAE by 41.8%. Detailed results can be found in figure (3).

In figure (4.a), we show the scatter plot of true price versus predicted price. We can see from the linear trend with positive slope that the predicted prices are highly correlated with true prices. In figure (4.b), we show the scatter plot of true price versus absolute error in price prediction. From the linear trend with positive slope, we can see that costlier products suffer a larger mean absolute error as expected.

4. Future Work

We see the following avenues of future work:

- Training a CNN for the task of price prediction instead of using a pretrained CNN for object classification.
- Comparing the algorithmic prediction performance to the performance of crowdsourced predictions from human annotators.
- Evaluating predictive power and statistical significance of visual features on product categories other than "Cellphones and Accessories."
- Investigating any heterogeneity in the predictive power and statistical significance of visual features across product categories.
- Combining multiple images provided for price prediction using CNN feature combining scheme like VLAD [5].
- Retrieving exact product from a product catalogue using pictures and simple metadata like product categories.

5. Conclusion

In this paper, we established that visual features are statistically significant in predicting product prices. We compared various regression algorithms which used visual features and product metadata to predict product prices. We found that Random Forest regression provides the best results. We intend to carry out steps outlined in the "Future Work" section to gain a better understanding of how a product's images can help decide the price that a seller can obtain for his product in online marketplaces like Amazon, Alibaba, or Flipkart.

References

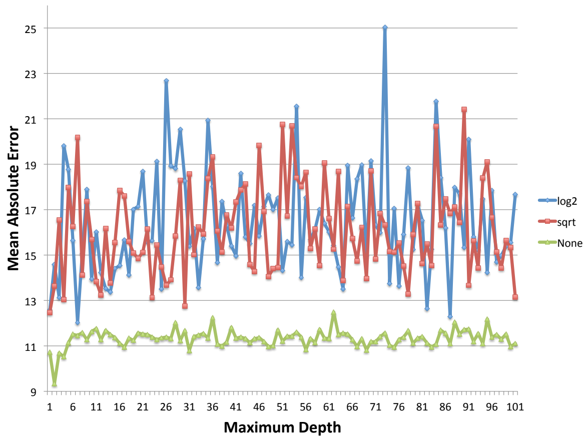
[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[2] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52. ACM, 2015.

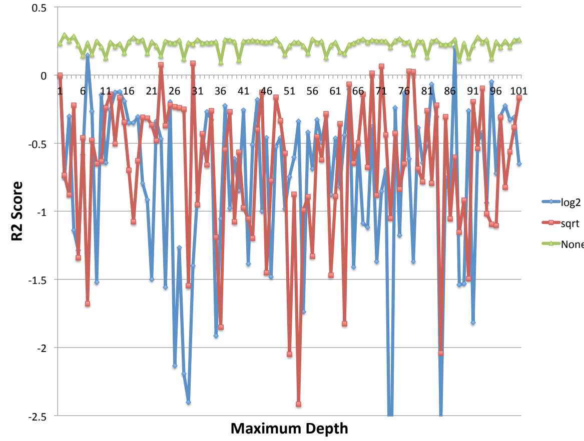
[3] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, pages 507–517. International World Wide Web Conferences Steering Committee, 2016.

[4] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

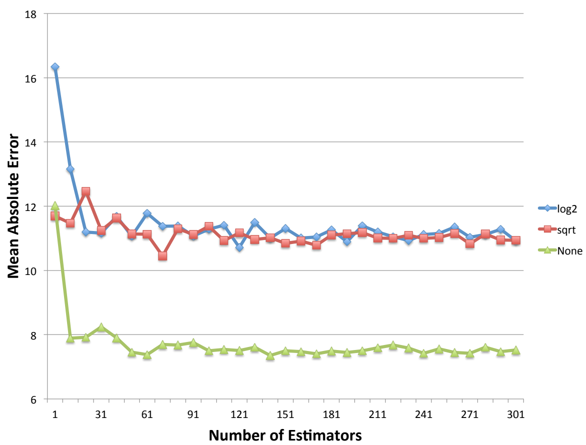
[5] Relja Arandjelovic and Andrew Zisserman. All about vlad. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1578–1585, 2013.



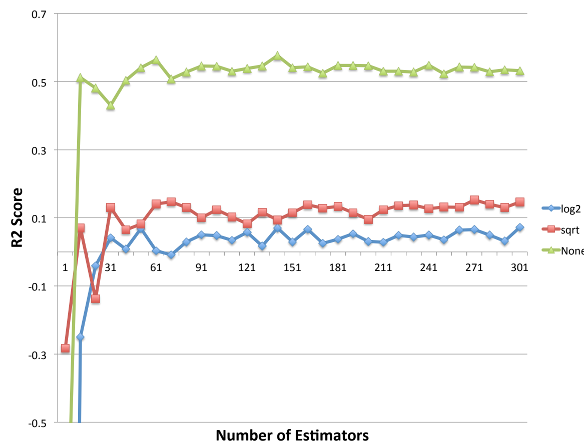
(a) MAE for Decision Tree



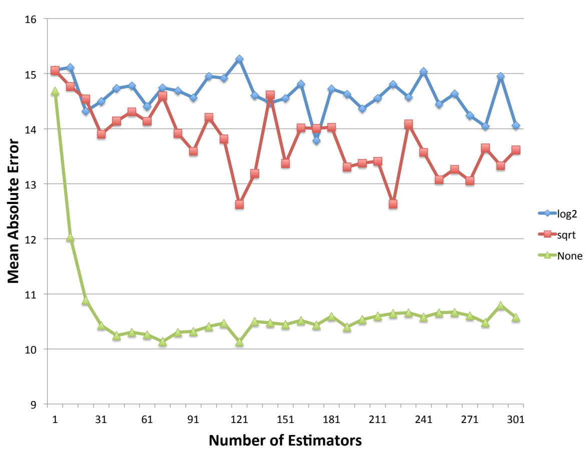
(b) R^2 for Decision Tree



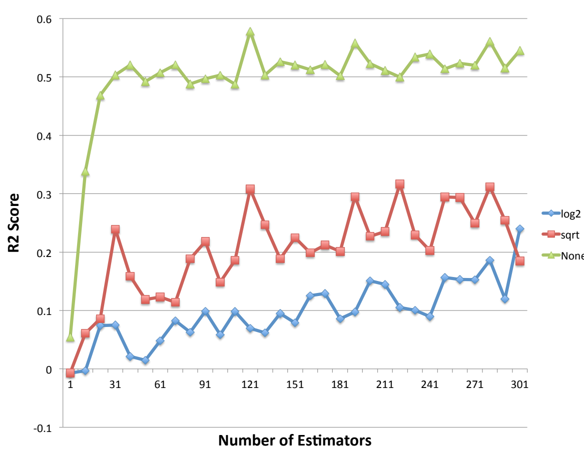
(a) MAE for Random Forest



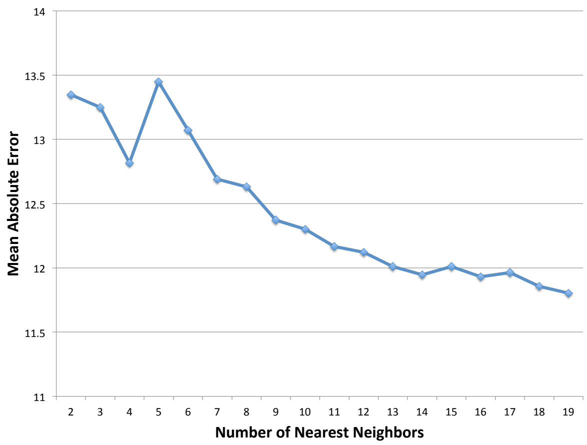
(b) R^2 for Random Forest



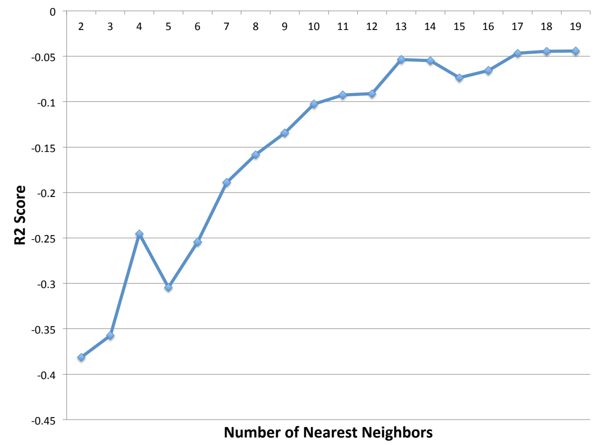
(a) MAE for Gradient Boosting Regression



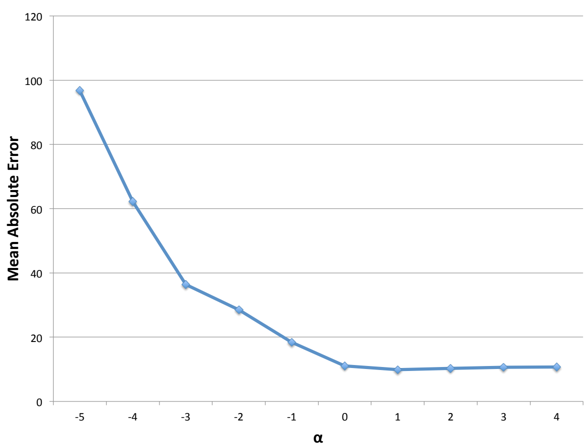
(b) R^2 for Gradient Boosting Regression



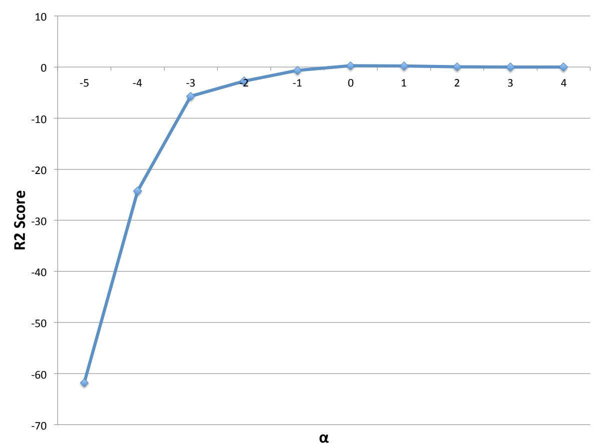
(a) MAE for KNN Regression



(b) R^2 for KNN Regression



(a) MAE for Ridge Regression



(b) R^2 for Ridge Regression

Table 3: *Regression results for the five algorithms. The figures on the left show Mean Absolute Error (MAE), and the figures on the right show R^2 score.*