CHAPTER 5
MEASURING PERFORMANCE IN REGRESSION MODELS

APPLIED PREDICTIVE MODELING BY KUHN & JOHNSON
COLLATED BY PROF. CHING-SHIH (VINCE) TSOU (PH.D.)
CENTER FOR APPLICATIONS OF DATA SCIENCE (CADS)
GRADUATE INSTITUTE OF INFORMATION AND DECISION SCIENCES (GIIDS)
NATIONAL TAIPEI UNIVERSITY OF BUSINESS (NTUB)
CHINESE ACADEMY OF R SOFTWARE (CARS)
DATA SCIENCE & BUSINESS APPLICATIONS (DSBA) ASSOCIATION OF TAIWAN

AGENDA

- Introduction
- Quantitative Measures of Performance
- The Variance-Bias Trade-off

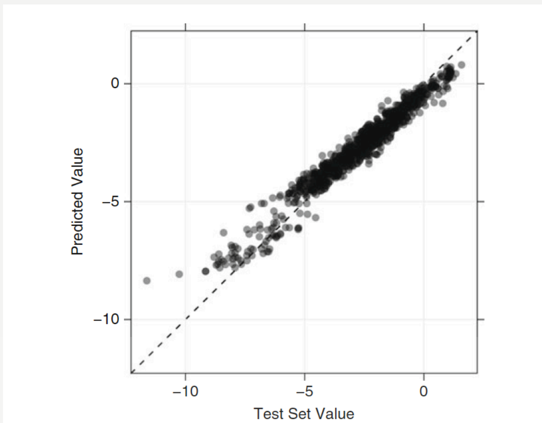
INTRODUCTION

- For models predicting a numeric outcome, some measure of accuracy is typically used to evaluate the effectiveness of the model.
- To understand the strengths and weaknesses of a particular model, relying solely on a single metric is problematic. (You had better try different performance measures to evaluate your models.)
- Visualizations of the model fit, particularly residual plots, are critical to understanding whether the model is fit for purpose.

QUANTITATIVE MEASURES OF PERFORMANCE – RMSE FOR ACCURACY

- When the outcome is a number, the most common method for characterizing a model's predictive capabilities is to use the RMSE (Root Mean Sqaured Error).
- The value is usually interpreted as either how far (on average) the residuals are from zero or as the average distance between the observed values and the model predictions.
- An R^2 value of 0.75 implies that the model can explain three-quarters of the variation in the outcome. There are multiple formulas for calculating this quantity (Kv˚alseth 1985), although the simplest version finds the correlation coefficient between the observed and predicted values (usually denoted by R) and squares it.

QUANTITATIVE MEASURES OF PERFORMANCE – R SQUARE FOR CORRELATION



- While this is an easily interpretable statistic, the practitioner must remember that R^2 is a measure of correlation, not accuracy.
- This figure shows an example where the R^2 between the observed and predicted values is high (51%), but the model has a tendency to overpredict low values and underpredict high ones.
- This phenomenon can be common to some of the tree-based regression models discussed in Chap. 8.

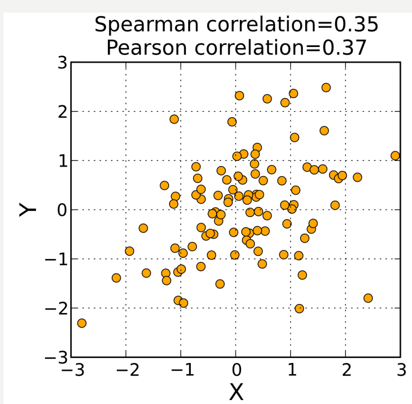
QUANTITATIVE MEASURES OF PERFORMANCE – CORRELATION VERSUS ACCURACY

- It is also important to realize that R2 is dependent on the *variation in the outcome*. (Wow ! What a good comment.)
- Using the interpretation that this statistic measures the proportion of variance explained by the model, one must remember that the denominator of that proportion is calculated using the sample variance of the outcome.
- **For example, suppose a test set outcome has a variance of 4.2. If the RMSE of a predictive model were 1, the R2 would be roughly 76 % ((4.2-1)/4.2 = 76.19%). If we had another test set with exactly the same RMSE, but the test outcomes were less variable, the results would look worse. For example, if the test set variance were 3, the R2 would be 67 % ((3-1)/3 = 66.67%).**
- Practically speaking, this dependence on the outcome variance can also have a drastic effect on how the model is viewed.
- For example, suppose we were building a model to predict the sale price of houses using predictors such as house characteristics (e.g., square footage, number of bedrooms, number of bathrooms), as well as lot size and location. If the range of the houses in the test set was large, say from $60K to $2M, the variance of the sale price would also be very large. One might view a model with a 90 % R2 positively, but the RMSE may be in the tens of thousands of dollars—poor predictive accuracy for anyone selling a moderately priced property.
- MEASURE of CORRELATION, NOT MEASURE of ACCURACY !

QUANTITATIVE MEASURES OF PERFORMANCE – PEARSON VERSUS SPEARMAN CORRELATION

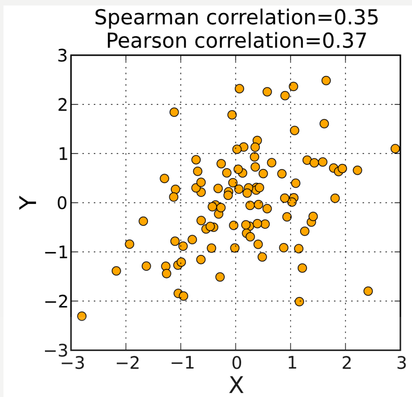
- In some cases, the goal of the model is to simply rank new samples. As previously discussed, pharmaceutical scientists may screen large numbers of compounds for their activity in an effort to find "hits."
- Here, the focus is on the ranking ability of the model rather than its predictive accuracy. (Again ! A different view of evaluation.)
- In this situation, determining the rank correlation between the observed and predicted values might be a more appropriate metric.
- The rank correlation takes the ranks of the observed outcome values (as opposed to their actual numbers) and evaluates how close these are to ranks of the model predictions.
- To calculate this value, the ranks of the observed and predicted outcomes are obtained and the correlation coefficient between these ranks is calculated.
- This metric is commonly known as Spearman's rank correlation which will be explain in next slide.

SPEARMAN'S RANK CORRELATION



- The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables.
- While Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not).
- If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

SPEARMAN'S RANK CORRELATION

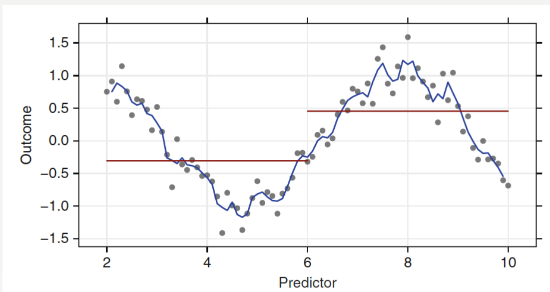


- Spearman's coefficient is appropriate for both continuous and discrete variables, including ordinal variables.
- Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient.

THE VARIANCE-BIAS TRADE-OFF – DECOMPOSITION OF MSE

- The MSE can be decomposed into more specific pieces. Formally, the MSE of a model is
  $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$
  - where y_i is the outcome and ˆy_i(y_i hat) is the model prediction of that sample's outcome.
- If we assume that the data points are statistically independent and that the residuals have a theoretical mean of zero and a constant variance of σ^2, then
  $E[\text{MSE}] = \sigma^2 + (\text{Model Bias})^2 + \text{Model Variance}$ where E is the expected value.
- The first part (σ2) is usually called "irreducible noise" and cannot be eliminated by modeling.
- The second term is the squared bias of the model. This reflects how close the functional form of the model can get to the true relationship between the predictors and the outcome.
- The last term is the model variance.

THE VARIANCE-BIAS TRADE-OFF – DEMO BY SINE WAVE



- This figure shows extreme examples of models that are either high bias or high variance.
- The model fit shown in red splits the data in half and predicts each half with a simple average.
- This model has low variance since it would not substantially change if another set of data points were generated the same way.
- The blue line is a three-point moving average. It is flexible enough to model the *sin* wave, but small perturbations in the data will significantly change the model fit. Because of this, it has high variance.

THE VARIANCE-BIAS TRADE-OFF – OVERFITTING VERSUS UNDERFITTING

- It is generally true that more complex models can have very high variance, which leads to over-fitting.
- On the other hand, simple models tend not to over-fit, but under-fit if they are not flexible enough to model the true relationship (thus high bias).
- Highly correlated predictors can lead to collinearity issues and this can greatly increase the model variance.
- In subsequent chapters, models will be discussed that can increase the bias (to some limit) in the model to greatly reduce the model variance as a way to mitigate the problem of collinearity. This is referred to as the *variance-bias trade-off*.


THANK YOU