

Evaluating Statistical Models for Stakeholder Needs

Mona Khalil
ODSC East
March 30, 2021

Getting Started

1. Download or fork the repository for this session:
github.com/mona-kay/odsc-model-evaluation
2. Open jupyter lab or jupyter notebooks
3. Choose an Exercise notebook based on your dataset of interest for the tutorial
 - a. Note: I will switch between slides and the Model Evaluation notebook

Overview

- What is model evaluation?
- Model evaluation metrics
 - Regressors*
 - Classifiers
- Modeling purposes
 - Models for explanation*
 - Models for prediction
- Key metrics of focus based on stakeholder needs
 - Understanding needs from questions and requests*
 - Selecting and evaluating models based on needs*

Pre-Requisites

- Familiarity with pandas and scikit-learn
- Jupyter Notebooks/Jupyter Lab installed
- Understanding of/experience with linear regression

About

Senior Data Scientist @ Greenhouse

- Teaching Python, SQL, and stats for >4 years
- **AI+ Python Fundamentals and SQL**
Bootcamp Instructor
- Passionate about responsible and effective analytics & methodological practices



Twitter: [mona_kay_](#)

What is model evaluation?

Model evaluation is part of the model development process in which one or more methods/metrics are used to determine the best possible fit model.

You create, develop and iterate on a model by...

- Generating features
- Fitting multiple algorithms
- Tuning hyperparameters

Model metrics allow you to **compare** versions of the models you create.

Pre-Exercise Exploration (~5 minutes)

Take a few minutes to explore the dataset.

Model Evaluation Metrics

Regression

R2 score - the amount of the variation in y which is predictable by the independent variable(s)

Mean squared error (MSE) - loss function measuring the average squared difference between estimated and true values

Classification

Accuracy score(s) - overall accuracy of classifying data points, balanced accuracy, true positive, true negative rates, etc.

Area under the curve (AUC) - the probability that a random positive sample will have a higher score than a random negative sample

Exercise 1 (~8 minutes)

Compare at least two models using scikit-learn.

- Take note of the **attributes** available with each model. What information does each provide you about the fit of the model?
- Compare the model output using at least two model metrics.

Evaluating Individual Features/Predictors

Linear Models

coefficients - indicates the size and directionality of a single unit of change in y

t-values and p-values for coefficients - indicators of the strength of the *partial correlational relationship* between the predictor and the outcome

Tree-based Models

feature importances - a relative metric indicating the importance of a feature when fitting the model/making a prediction

- Used to rank predictors in order of importance
- *Directionality* of features does not apply here

Modeling Purposes

- Modeling for explanation
 - Explain the fit and shape of predictors and their relationship with an outcome
 - Answering questions about actions already taken
 - What recommendations should be made to diabetes patients to improve their outcomes?
 - A series of drugs to improve blood glucose levels was distributed to diabetic patients. Was it effective?
- Modeling for prediction
 - Generate predictions as actionable insights
 - Answering questions *before* an outcome has been reached
 - What is the expected disease progression of diabetes for a series of 1,000 newly diagnosed patients?
 - Which customers are predicted to cancel their contracts in the next year?

Modeling Strategies by Purpose

- Modeling for explanation
 - Leverage tools that allow for evaluation of individual predictors
 - statsmodels provides a detailed summary
 - Optimize for key relationships between predictors and outcomes
- Modeling for prediction
 - Leverage tools that allow for ease of tuning/iteration
 - scikit-learn is optimized for this
 - Optimize for overall accuracy of the model

Exercise 2 (~5 minutes)

Explore the statsmodels OLS regression summary.

- What features are not statistically significant predictors of y in this model? Try removing them (drop them and recreate the training/test set) and re-evaluate the strength and directionality of the t-values.

Addressing Stakeholder Needs

- Explanation
 - Model Selection
 - Use models and tools that include the strength and directionality of relationships
 - Use models that allow for comprehensive reporting on relationships
 - Optimize for the *strength* and *shape* of relationships with key predictors and the outcome
 - Reporting on Results
 - Provide a written summary of conclusions, including

Exercise 3 (~8 minutes)

- Prediction
 - Model selection
 - Iterate over several models, optimizing for overall performance metrics
 - Carefully select the most important performance metric based on the stakeholder question
 - Providing results
 - Identify the appropriate format for sharing results with stakeholders
 - Determine what steps are necessary for the model to be appropriately leveraged