ML_Chap.2Notes

Summary/Review - Week1 (2021.12)

Introduction to Supervised Machine Learning

The types of supervised Machine Learning are:

- **Regression**, in which the target variable is continuous
- **Classification**, in which the target variable is categorical

To build a classification model you need:

- Features that can be quantified
- A labeled target or outcome variable
- Method to measure similarity

Linear Regression

A linear regression models the relationship between a continuous variable and one or more scaled variables.It is usually represented as a dependent function equal to the sum of a coefficient plus scaling factors times the independent variables.

Residuals are defined as the difference between an actual value and a predicted value.

A modeling best practice for linear regression is:

- Use cost function to fit the linear regression model
- Develop multiple models
- Compare the results and choose the one that fits your data and whether you are using your model for prediction or interpretation.

Three common **measures of error** for linear regressions are:
- Sum of squared Error (SSE)
- Total Sum of Squares (TSS)
- Coefficient of Determination (R2)

Linear Regression Syntax

The most simple syntax to train a linear regression using scikit learn is:

from sklearn.linear_model import LinearRegression
LR = LinearRegression()
LR = LR.fit(X_train, y_train)

To score a data frame X_test you would use this syntax:
y_predict = LR.predict(X_test)

Summary/Review - Week2

Training and Test Splits

Splitting your data into a training and a test set can help you choose a model that has better chances at generalizing and is not overfitted.

The training data is used to fit the model, while the test data is used to measure error and performance.

Training error tends to decrease with a more complex model.
Cross validation error generally has a u-shape.
It decreases with more complex models, up to a point in which it starts to increase again.

Cross Validation

The three most common cross validation approaches are:
- k-fold cross validation
- leave one out cross validation
- stratified cross validation

Polynomial Regression

Polynomial terms help you capture nonlinear effects of your features.

Other algorithms that help you extend your linear models are:
- Logistic Regression
- K-Nearest Neighbors
- Decision Trees
- Support Vector Machines
- Random Forests
- Ensemble Methods
- Deep Learning Approaches

Summary/Review - Week3

Regularization Techniques (Important key of ML)

Three sources of error for your model are:

bias, variance, and, irreducible error.

Regularization is a way to achieve building simple models with relatively low error.
It helps you avoid overfitting by penalizing high-valued coefficients.
It reduces parameters and shrinks the model.

Regularization adds an adjustable regularization strength parameter directly into the cost function.
Regularization performs feature selection by shrinking the contribution of features, which can prevent overfitting.

In Ridge Regression, the complexity penalty λ is applied proportionally to squared coefficient values.
(L2 regularization penalty)
– The penalty term has the effect of "shrinking" coefficients toward 0.
– This imposes bias on the model, but also reduces variance.
– We can select the best regularization strength lambda via cross-validation.
– It's a best practice to scale features (i.e. using StandardScaler) so penalties aren't impacted by variable scale.

In LASSO regression: the complexity penalty λ (lambda) is proportional to the absolute value of coefficients.
(L1 regularization penalty)
LASSO stands for : Least Absolute Shrinkage and Selection Operator.

– Similar effect to Ridge in terms of complexity tradeoff: increasing lambda raises bias but lowers variance.
– LASSO is more likely than Ridge to perform feature selection, in that for a fixed λ, LASSO is more likely to result in coefficients being set to zero.

Elastic Net combines penalties from both Ridge and LASSO regression.
It requires tuning of an additional parameter that determines emphasis of L1 vs. L2 regularization penalties.

LASSO's feature selection property yields an interpretability advantage, but may underperform if the target truly depends on many of the features.

Elastic Net, an alternative hybrid approach, introduces a new parameter α (alpha) that determines a weighted average of L1 and L2 penalties.

Regularization techniques have an analytical, a geometric, and a probabilistic interpretation.