Notation

Machine Learning - Section 1.1.1

Marc Omar Haddad

17 January, 2020

In machine learning, data comes in the form of the **outcome** we wish to predict, and the **features** to be used to predict the outcome. Our goal: Build an algorithm that takes features as inputs, and returns a prediction for the unknown outcome. Specifically, Machine Learning involves the use of a dataset of *known* outcomes to *train* our algorithm to predict the outcomes of similar datasets with *unknown* outcomes. Y is used to denote outcomes; and $X_1, ...X_p$ is used to denote the various features. Features are also referred to as **predictors** and **covariates**.

Prediction problems are divided into **Categorical** outcomes and **Continuous** outcomes.
For Categorical outcomes, Y can be any of K classes: $Y_1, ..., Y_K$. For example, an algorithm that is trained to read digits has $K = 10$, with the classes being $[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]$. In speech recognition, the outcome is all of the possible words we are trying to detect. Spam detection has two outcomes, spam or not spam. Thus, for binary data we use $k = 0, 1$, since there are only 2 outcomes.

General Setup: We have a series of features and an unknown outcome we wish to predict.

outcome	feature_1	feature_2	feature_3	feature_4	feature_5
?	X_1	X_2	X_3	X_4	X_5

To build an algorithm that can predict the above unknown outcome we collect data for which we *do* know the outcome.

outcome	feature_1	feature_2	feature_3	feature_4	feature_5
Y_1	X_1,1	X_2,1	X_3,1	X_4,1	X_5,1
Y_2	X_1,2	X_2,2	X_3,2	X_4,2	X_5,2
Y_3	X_1,3	X_2,3	X_3,3	X_4,3	X_5,3
Y_4	X_1,4	X_2,4	X_3,4	X_4,4	X_5,4
Y_5	X_1,5	X_2,5	X_3,5	X_4,5	X_5,5

We use $\hat{Y}$ to denote **prediction**. The term "**actual outcome**" is used to denote what we ended up *actually* observing. Thus, we want the prediction $\hat{Y}$ to match the actual outcome Y. Y can be categorical (spam/not spam, digits [0-9], letters [A-Z], etc.) or continuous (ratings, prices, profit, etc.).
When the outcome is categorical, we refer to our Machine Learning task as **Classification**. Our predictions will be categorical, just like our outcomes, and will either be *correct* or *incorrect*. When the outcome is continuous, the Machine Learning task is referred to as a **Prediction**. With predictions there is no "right" or "wrong" answers; Due to their continuous nature, we measure the **error** when assessing predictive

algorithms. The error is simply the difference between the prediction and the actual outcome.