Reflection Week 1

Summary

In this week we have covered on the overview of the machine learning world, as well as methodology of assessing machine learning models and concepts used in the assessment of these models. We also learn about regularization, regressions and scaling, which help to optimize our models' efficiency.

Concepts

1. Regression: Regression is a type of model that fits numerical data to numerical data. Regression models often fit into equations that machine learning algorithms learned to form a generalized pattern of the data.
2. Parameters, hyperparameters and variables:

parameter	hyperparameter	variable
quantities defining a theoretical model that are learned during training	explicitedly configuration values that are set beforehand	quantities varying from individual to individual

3. Residual Sum of Squares: $RSS = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2$ is a quality metric used to evaluate models.
4. Gradient Descent is an optimization algorithm commonly used to train models. From a starting point that helps us to evaluate the performance, we can find the derivative (slope) to observe which way to go down to the convergence point. The convergence point is the lowest point on the curve, which represents the lowest cost function value. Based on the slope value, we can adjust the parameters (weight and bias) and go down the slope, then continue calculating the slopes until we hit the convergence point.
5. True error is the error that happened when we perform the model on untouched data. True error may decrease when we increase the complexity of our models, but increases when we overfit (fail to generalize)
6. Train error is the error that happened when we trained the model. Train error decreases as we increase the complexity of the model.
7. Bias-Variance trade-off: bias is a type of error that happens when we fail to fit the signal (underfit), and variance errors result when the model memorizes answers rather than learning the general patterns. When there is high bias, there is low variance and vice versa. $Error = Bias^2 + Variance + Noise$
8. Two ways to choose complexity of a training model without ruining the test set (which should be untouched by the training model)

method	validation set	cross-validation
description	divide train dataset to a smaller train dataset and a validation set, then use the validation set to choose the complexity	divide train dataset into smaller chunks of sets, then use one chunk to validate and the rest to train and repeat with other chunks
pro	save time, easy to implement	don't have to get rid of any training data
con	sacrifice training data	harder to implement, needs more time

9. Ridge Regression $RSS(W) + \lambda||x||_2^2$, with λ as a tuning parameter that changes how much the model cares about the regularization term. When this parameter = 0, then the estimates produced by the regression will be equal to least squares; but when it goes to infinity, the hence the ridge regression coefficient estimates will approach zero, because we want to minimize this regression as much as we can, and since the impact of the lambda is immense, we have no choice but to shrink down the coefficients so that the RSS part is as small as possible to minimize the cost function.
10. Regularization is a form of regression that constrains the coefficient estimates toward zeros, so that it discourages learning a more complex model to avoid the risk of overfitting. Using RSS
11. Scaling helps to generalize data points so that the features in the dataset are in the same footing without any upfront importance (like one feature has significant larger values because it is in a different scale). Scaling also helps to make these features unitless, bring the data points closer together, and speed up training.

Uncertainties

1. How does regularization work mathematically?
2. Why do the true error and train error reach an equilibrium but never cross as the train set increases in size?