WKS5 – Error-based Learning

Theory

Review the videos from week 5. Answer the following questions based on those lectures

1. Explain the difference between classification and regression

Answer: Unlike classification, which makes predictions from among a finite range of categorical values, regression models output continuous values as predictions of some target feature values

1. Describe the basic idea behind error-based learning

Answer: An error-based learning approach uses systematic feedback techniques to internalise the strategy of stopping, checking, and correcting one's own errors during task performance.

1. What is the predictor function in linear regression

Answer: For a two-feature case when there is a linear relationship, we can fit a line that relates the features as a linear equation of a y-axis intercept b and a slope m. The formula of the straight line is the predictor function in linear regression.

1. What is the role of the error function in determining the optimal predictor function?

Answer: measures the total residuals from the line to the points in our dataset.

1. What is gradient decent and, briefly, how does it work?

Answer: Gradient decent is the way we will automatically find this minimum. The algorithm starts with some randomly initialised points on the hyperplane and the proceeds to find the global minimum with respect to all the dimensions until it converges at the lowest point.

1. How should be treat categorical descriptive features in linear regression?

Answer: Categorical descriptive features need to be transformed before including in a linear model. This is because the arbitrary numerical values in the feature range are meaningless as multipliers of weights for a model fit. One simple and popular way to transform categorical features is called one hot encoding. In this process we extend the dataset by adding one new column for every possible value of the categorical feature. Then we put a one in that new column for the corresponding feature value in the original descriptive feature. During model training, we exclude the old column and only consider the newly created ones.

Practice

Follow the tutorial videos from week 5 and carry out the following steps

1. Download the code archive and extract the file from the week 4 learning materials. Make sure that you can run the examples code as provided.
2. Expand the **sse.ipynb** notebook provided to plot the error function for the w[0] weight. Verify that the local minimum, the derivate of this cost function, is zero as expected
3. Add a 3D error surface of the w[0] and w[1] parameters from this notebook.
4. In the **linear.ipynb** notebook, add error function plots for the “bmi” feature of the dataset with respect to w[0] and w[1]. Find estimates for parameters from the plots and draw the best fit line on the scatter graph
5. Write a function with implements one-hot encoding for a categorical feature of an arbitrary size
6. Implement a basic version of the gradient descent algorithm in Python using the **loss()** function presented in **sse.py** as a starting point. Test your solution on a randomly generated regression dataset and verify your results match those of the corresponding error function minimums (from the plots)