Model solutions
INF264

23.11.2020

1 Explanations

1)

While making wrong assumptions can certainly hurt, you cannot generalise without making assumptions, that is, *inductive bias*.

2)

Alice has an underfitting model (high bias), Bob overfits (high variance). We can use the *bias-variance tradeoff* to analyse the relationship between model complexity and different errors.

3)

Bob is wrong. Based on the *no free lunch theorem* there is no such thing as the best machine learning algorithm.

4)

The problem is that k-means clustering finds clusters that are "spherical" and the clusters in Alice's data are not.

5)

The data is imbalanced and thus accuracy is not a good performance measure. Bob's classifier predicts always that the component works so it is useless.

2 Dimensionality reduction

Errors: In feature selection, we select a subset of *features*.

In PCA, the first principal component is the direction that maximizes variance. That is, the red arrow in the figure.

3 Neural networks

Errors: By definition, $relu(z) = \max(0, z)$.

Neural networks can be used in unsupervised learning. For example, aurtoencoders for dimensionality reduction.

Neural networks with only one hidden layer are universal approximators, that is, they can approximate continuous functions up to arbitrary precision (given enough neurons).

Having lots of parameters makes a model more likely to overfit.

4 Gradient descent

$$E[w, b] = \sum_{i=1}^{n} (y_i - wx_i - b)^4$$

Partial derivatives:

$$\frac{\partial E}{\partial w} = -\sum_{i=1}^{n} 4x_i(y_i - wx_i - b)^3$$

$$\frac{\partial E}{\partial b} = -\sum_{i=1}^{n} 4(y_i - wx_i - b)^3$$

Now the updates are

$$w_1 \leftarrow w_0 + \gamma \sum_{i=1}^{n} 4x_i(y_i - w_0x_i - b_0)^3$$

and

$$b_1 \leftarrow b_0 + \gamma \sum_{i=1}^{n} 4(y_i - w_0x_i - b_0)^3$$

Plug-in the values and we get
Loss before the update=0.0272
Updated parameters: $w_1 = 1.16$, $b_1 = 0.79$
Loss after the update=0.005

5 Hierarchical clustering

$$\Big(\big(C, (A, F)_2\big)_3, \Big(E, (B, D)_1\Big)_4\Big)_5$$

6 Model selection

1. Model overfits if training accuracy/F1 is significantly larger than validation accuracy/F1.

2. Choose the model with largest validation accuracy/F1

3. Based on the information that we have, we cannot compute an unbiased estimate on the performance of the selected on unseen data (We would need test data to do that).