ML - Tips and Tricks

R Basheer Ahammad

1 Diagnostics

In the context of machine learning, two important concepts that influence the performance of predictive models are bias and variance.

1.1 Bias

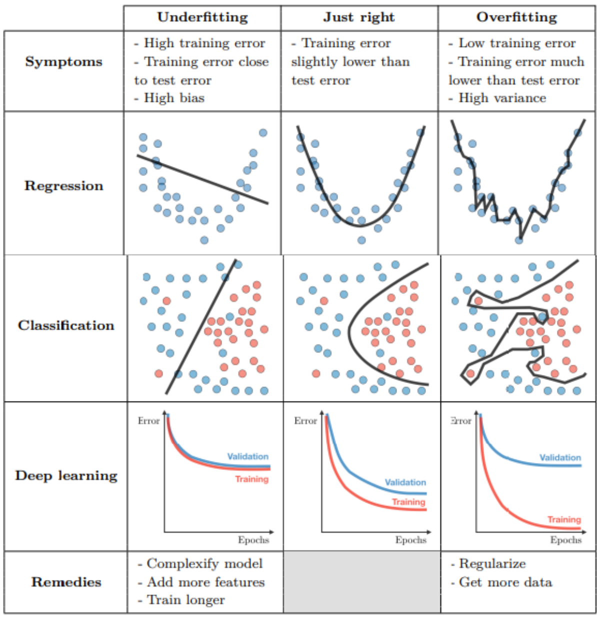
The bias of a model is the difference between the expected prediction and the correct model that we try to predict for given data points.

1.2 Variance

The variance of a model is the variability of the model prediction for given datapoints.

1.3 Bias-Variance Trade-off

The simpler the model, the higher the bias, and the more complex the model, the higher the variance.



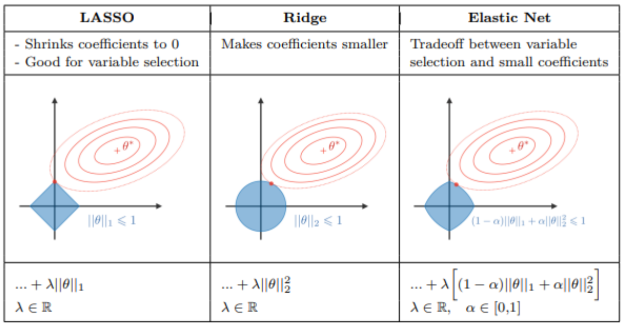
	Underfitting	Just right	Overfitting
Symptoms	- High training error - Training error close to test error - High bias	- Training error slightly lower than test error	- Low training error - Training error much lower than test error - High variance
Regression			
Classification			
Deep learning			
Remedies	- Complexify model - Add more features - Train longer		- Regularize - Get more data

1.4 Remedies

- For Underfitting :
  - Complexify model
  - Add more features
  - Train longer
- For Overfitting :
  - Regularize
  - Get more data

2 Regularization

The regularization procedure aims at avoiding the model to overfit the data and thus deals with high variance issues. The following table sums up the different types of commonly used regularization techniques:



LASSO	Ridge	Elastic Net
- Shrinks coefficients to 0 - Good for variable selection	Makes coefficients smaller	Tradeoff between variable selection and small coefficients
$... + \lambda\|\|\theta\|\|_1$ $\lambda \in \mathbb{R}$	$... + \lambda\|\|\theta\|\|_2^2$ $\lambda \in \mathbb{R}$	$... + \lambda\Big[(1-\alpha)\|\|\theta\|\|_1 + \alpha\|\|\theta\|\|_2^2\Big]$ $\lambda \in \mathbb{R}, \quad \alpha \in [0,1]$

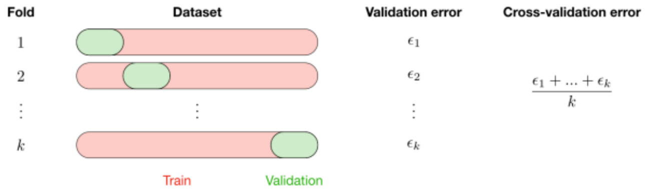
$+\theta^*$
$||\theta||_1 \leqslant 1$
$+\theta^*$
$||\theta||_2 \leqslant 1$
$+\theta^*$
$(1-\alpha)||\theta||_1 + \alpha||\theta||_2^2 \leqslant 1$

3 Cross Validation

Cross-validation, also noted CV, is a method that is used to select a model that does not rely too much on the initial training set. The different types are summed up in the table below:

k-fold	Leave-p-out
- Training on $k-1$ folds and assessment on the remaining one - Generally $k=5$ or 10	- Training on $n-p$ observations and assessment on the p remaining ones - Case $p=1$ is called leave-one-out

The most commonly used method is called **k-fold cross-validation**, which splits the training data into k folds. During each iteration, the model is validated on one fold while being trained on the remaining $k-1$ folds. This process is repeated k times, ensuring that each fold is used as a validation set exactly once. The errors obtained from these iterations are averaged over the k folds to calculate the **cross-validation error**.



Fold
Dataset
Validation error
Cross-validation error
1
2
⋮
k
⋮
ϵ_1
ϵ_2
⋮
ϵ_k
$\frac{\epsilon_1 + ... + \epsilon_k}{k}$
Train
Validation