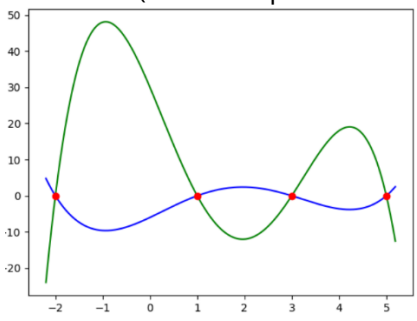
Overfitting

1. What it means –
    a. Model does well on training data and not on validation/test data.
    b. Train loss too low and val loss high (Good model should have both values ~ same)
    c. Model learns training data too well, poor generalization on test data
    d. Model learns desired function as well as noise
    e. Coefficients are inflated (As size of parameters increase, overfitting increase. See fig for example: 

Green curve: $h_1(x) = -x^4 + 7x^3 - 5x^2 - 31x + 30$;
Blue curve: $h_2(x) = \frac{x^4}{5} - \frac{7x^3}{5} + x^2 + \frac{31x}{5} - 6$
    f. More common in non-parametric methods like decision trees
2. Why it happens –
    a. Number of Parameters > Number of training samples (i.e. too complex model)
    b. Number of features are too large
    c. Training data is imbalanced
    d. Model is trained on the data for too long, it starts learning noise
    e. Too long training reduce bias but increase variance
    f. Training data is too small
3. How to detect –
    a. Cross-validation
    b. Train loss, val loss
4. How to avoid Overfitting –
    a. Resampling
    b. Dimensionality reduction i.e. reduce number of features
    c. Choose a simpler model (or not make model too complex)
    d. Early stopping
    e. Ensemble methods – Bagging & Boosting
    f. Data Augmentation
    g. Pruning – feature selection
    h. Feature Normalization
    i. Regularization
        1. L1 or LASSO (Least Absolute Shrinkage and Selection Operator)
        2. L2 or Ridge

            3. Elastic net
    j. Dropout
        i. How it works
            1. Randomly deactivates a fraction of neurons in a layer
        ii. Why it works
            1. It simulates training multiple neural networks in parallel (similarity with RF), forcing it to be more robust and preventing it from relying too much on one neuron
5. Difference between overfitting and underfitting?
    a. Underfitting
        i. Model is too simple to learn patterns
        ii. High Bias
        iii. Performs poorly on both train and test data
        iv. Could be caused due to improper scaling of features
    b. Overfitting
        i. Model is too complex that it starts learning noise
        ii. High Variance
        iii. Performs well on train data and poorly on test data
6. Methods to avoid underfitting or high bias
    a. Complex model
    b. Increase num of features
    c. Normalize features
    d. Feature Engineering