Springboard - Blog

https://www.springboard.com/blog/machine-learning-interview-questions/

1. Bias vs Variance
    - **Bias** is due to erroneus or overly simplistic assumptions in learning algorithm
    - Usually underfitting your data
    - **Variance** typically due to too much complexity in learning algorithm
    - Makes the model sensitive to high degrees of variation in training data.
    - Too much noise from training data
    - If you make data more complex and add more variables, you'll lose bias but gain variance.
2. Supervised vs Unsupervised learning
    - Supervised requires labeled data. Unsupervised does not.
3. How is KNN different from k-means clustering?
    - KNN is a supervised classification algorithm.
    - K-means clustering is unsupervised.
    - Works very similarly
    - KNN required labelled data
    - K means clustering requires only a set of unlabeled point and a threshold
    - The algorithm will gradually *learn* how to cluster them by computing mean of the distance between different points.
4. How does a ROC curve work
    - graphical representation of constrast between true and false positive rate at various threholds.
    - Used as a proxy for trade-off between sensitivity of model (true positive) vs the fall-out or probability it will trigger a false alarm (false positives)
