

ML oral exam questions

- SVM algorithm (both parametric and not, even though afaik there is only the non parametric version in our slides). X3
- cost of decision trees (max depth) X4
- underfitting and overfitting in k-NN (size of k) X5
- pros and cons of oversampling and undersampling. X2
- describe entropy X4
- Information gain: equation and meaning when building a decision tree x2
- SVM X5
- meaning of C parameter
- regularization term meaning X3
- meaning of overfitting/underfitting X4
- oversampling and undersampling (pros and cons)
- when SMOTE fails. X2
- What is scaling and how is it implemented? X3
- Why is scaling useful for nn and k-nn?
- What does the fit method do?
- What are the definitions of parameter and hyperparameter?
- Perceptron X4
- Logistic Regression X3
- Linear Regression
- Difference between OVO and OVR X3
- decision tree X5
- Difference between precision and recall X3
- k-NN X2
- What is Bias and how can it be classified
- How does a neural network learn the best representation of the data
- Why SVM are slower than ANN x3
- (Comp Sci bachelors) Algorithmic runtime of several models: Decision tree = maxdepth, SVM, NN
- Why can gradient descent in NN get stuck on a local minimum, but not in linear regression (because the loss function of LR has a parabolic shape, so there's only one minimum; the loss f-n of a neural network can be a different shape and have many local minima)
- When do we stop when building a decision tree (when all training samples are classified)
- K-means X2
- Transform perceptron into its kernelized version

- A trick to transform the NN loss function into a convex function (it was very difficult question about the fact that I can exploit the cross entropy and calculate for pair wise layers to have a convex function each). x2
- KNN with $K = 1$, when I'm certain that I will always have overfitting when $K = 1$? With linearly separable data. He made an example of KNN with 700 positive labels and 300 negative labels.
- Closed form solution for linear regression
- How to train multiclass for knn and neural network
- Why svm is better than normal perceptron
- Regularization parameter, human word for the cost function equation. Why doing regularization parameter
- For tweet project, is there any method for vectorizing the tweet other than tokenizing the word (yes, word relation)
- Parametric vs non-parametric models x2
- Kernel methods, pros
- How to make NN less "complex"/overfit (early stopping)
- Generalized crossentropy loss function
- If we can process graphs with ML algorithm (subtree kernel function)
- What could be an alternative to GridSearchCross validation which cost less in term of computation time ? (answer = Hold-out, see slide lect. 7)
- "Are kernel methods just a fancy way to say we translate the samples in a higher dimensional space ?" (I'll let u think about this philosophical question)
- Multilabel classification
- OVO and OVR (in particular what happens if there is a tie between 2 classes)