# 机器学习必备资源 – 高频面试题大汇总

The Machine Learning part of the interview is usually the most elaborate one. That's the reason we have dedicated a complete post to the interview questions from ML. We've also provided, wherever possible, the link to Suggested Reading material that will be helpful in answering these questions.
We update these links from time to time and if you have any solution you can suggest please feel free to post it. You should explore these questions thoroughly, especially the ones that may relate to your previous experience and projects.

## General ML Questions

Here is a nice post covering various aspects of machine learning that'll be a good starting point.

- How will you differentiate a machine learning algorithm from other algorithms? Suggested reading

- What's the difference between data mining and machine learning?

- What are the advantages of machine learning?Often much more accurate than human-crafted rules (since data driven) · humans often incapable of expressing what they know (e.g., rules of English, or how to recognize letters), but can easily classify examples · don't need a human expert or programmer · cheap and flexible — can apply to any learning task

- Describe some popular machine learning methods. Suggested Reading

- How will you differentiate between supervised and unsupervised learning? Give few examples of algorithms for supervised learning? Suggested reading

- What is your favorite ML algorithm? How will you explain it to a layman? Why is it your favourite?

## Regression

- Is regression some type of supervised learning? Why?

- Explain the tradeoff between bias and variance in a regression problem.

- A learning algorithm with low bias and high variance may be suitable under what circumstances?

- What is regression analysis?

- What do coefficient estimates mean?

- How do you measure fit of the model? What do R and D mean?

- What are some possible problems with regression models? How do you avoid or compensate for them?

- Name a few types of regression you are familiar with? What are the differences?

- What are the downfalls of using too many or too few variables for performing regression?

# Linear Regression

Suggested reading on difference between linear and non-linear regression

- What is linear regression? Why is it called linear?

- What are the constraints you need to keep in mind when using a linear regression?

- How does the variance of the error term change with the number of predictors, in OLS?

- In linear regression, under what condition $R^2$ always equals a perfect 1?

- Do you consider the models Y~X1+X2+X1X2 and Y~X1+X2+X1X2 to be linear? Why? Suggested reading

- Do we always need the intercept term? When do we need it and when do we not? Suggested reading

- What is collinearity and what to do with it? Suggested reading

- How to remove multicollinearity? Suggested reading

- What is overfitting a regression model? What are ways to avoid it? Suggested reading

- What is Ridge Regression? How is it different from OLS Regression? Why do we need it?

- What is Lasso regression? How is it different from OLS and Ridge?

- What are the assumptions that standard linear regression models with standard estimation techniques make?

- How can some of these assumptions be relaxed?

2

- You fit a multiple regression to examine the effect of a particular feature. The feature comes back insignificant, but you believe it is significant. How will you explain it?

- Your model considers the feature X significant, and Z is not, but you expected the opposite result. How will you explain it?

- How to check if the regression model fits the data well?

- When to use k-Nearest Neighbors for regression?

- Could you explain some of the extension of linear models like Splines or LOESS/LOWESS?

# Classification

**Basic Questions**

- State some real life problems where classification algorithms can be used? Text categorization (e.g., spam filtering) · fraud detection · optical character recognition · machine vision (e.g., face detection) · natural-language processing (e.g., spoken language understanding) · market segmentation (e.g.: predict if customer will respond to promotion) · bioinformatics (e.g., classify proteins according to their function) etc.

- What is the simplest classification algorithm?Many consider Logistic Regression as a simple approach to begin with in order to to set a baseline and only make it more complicated if need be.

- What is your favourite ML algorithm? Why is it your favourite? How will you describe it to a non-technical person.

# Decision Trees

To answer questions on decision trees here are some useful links:
Youtube video tutorial
This article covers decision tree in depth
Other suggested reading

- What is a decision tree?

- What are some business reasons you might want to use a decision tree model?

- How do you build a decision tree model?

- What impurity measures do you know?

- Describe some of the different splitting rules used by different decision tree algorithms.

- Is a big brushy tree always good?

- How will you compare a decision tree to a logistic regression? Which is more suitable under different circumstances?

- What is pruning and why is it important?Ensemble models:To answer questions on ensemble models here is a useful link:

- Why do we combine multiple trees?

- What is Random Forest? Why would you prefer it to SVM?

Logistic regression:
Link to understand basics of Logistic regression
Here's a nice tutorial from Khan Academy

- What is logistic regression?

- How do we train a logistic regression model?

- How do we interpret its coefficients?

**Support Vector Machines**
A tutorial on SVM can be found here and here
- What is the maximal margin classifier? How this margin can be achieved and why is it beneficial?

- How do we train SVM? What about hard SVM and soft SVM?

- What is a kernel? Explain the Kernel trick

- Which kernels do you know? How to choose a kernel?

**Neural Networks**
Here's a link to Neural Network course from Hinton on Coursera
- What is an Artificial Neural Network?

- How to train an ANN? What is back propagation?

- How does a neural network with three layers (one input layer, one inner layer and one output layer) compare to a logistic regression?

- What is deep learning? What is CNN (Convolution Neural Network) or RNN (Recurrent Neural Network)?

Other models:
- What other models do you know?

- How can we use Naive Bayes classifier for categorical features? What if some features are numerical?

- Tradeoffs between different types of classification models. How to choose the best one?

- Compare logistic regression with decision trees and neural networks.

# Regularization

Suggested Reading: wikipedia and Quora answers
- What is Regularization?

- Which problem does Regularization try to solve?

Ans. used to address the overfitting problem, it penalizes your loss function by adding a multiple of an L1 (LASSO) or an L2 (Ridge) norm of your weights vector w (it is the vector of the learned parameters in your linear regression).

- What does it mean (practically) for a design matrix to be "ill-conditioned"?

- When might you want to use ridge regression instead of traditional linear regression?

- What is the difference between the L1 and L2 regularization?

- Why (geometrically) does LASSO produce solutions with zero-valued coefficients (as opposed to ridge)?

# Dimensionality Reduction

Suggested Reading: Scikit and Kdnuggets
- What is the purpose of dimensionality reduction and why do we need it?

- Are dimensionality reduction techniques supervised or not? Are all of them are (un)supervised?

- What ways of reducing dimensionality do you know?

- Is feature selection a dimensionality reduction technique?

- What is the difference between feature selection and feature extraction?

- Is it beneficial to perform dimensionality reduction before fitting an SVM? Why or why not?

## Principal Component Analysis

- What is Principal Component Analysis (PCA)? Under what conditions is PCA effective? How is it related to eigenvalue decomposition (EVD)?

- What are the differences between Factor Analysis and Principal Component Analysis?

- How will you use SVD to perform PCA? When SVD is better than EVD for PCA?

- Why do we need to center data for PCA and what can happen if we don't do it?

- Do we need to normalize data for PCA? Why?

- Is PCA a linear model or not? Why?

## Other Dimensionality Reduction techniques:

- Do you know other Dimensionality Reduction techniques?

- What is Independent Component Analysis (ICA)? What's the difference between ICA and PCA?

- Suppose you have a very sparse matrix where rows are highly dimensional. You project these rows on a random vector of relatively small dimensionality. Is it a valid dimensionality reduction technique or not?

- Have you heard of Kernel PCA or other non-linear dimensionality reduction techniques? What about LLE (Locally Linear Embedding) or tt-SNE (tt-distributed Stochastic Neighbor Embedding)

- What is Fisher Discriminant Analysis? How it is different from PCA? Is it supervised or not?

# Cluster Analysis

Suggested reading: tutorialspoint and Lecture notes

- Why do you need to use cluster analysis?

- Give examples of some cluster analysis methods?

- Differentiate between partitioning method and hierarchical methods.

- Explain K-Means and its objective?

- How do you select K for K-Means?

- How would you assess the quality of clustering?

# Optimization

Here is a good video to learn about optimization.

## Some basic questions about optimization

- Give examples of some convex and non-convex algorithms.

## Examples of convex optimisation problems in machine learning

linear regression/ Ridge regression, with Tikhonov regularisation, etc; sparse linear regression with L1 regularisation, such as lasso; Support vector machines; Parameter estimation in linear-Gaussian time series (Kalman filter and friends)

## Typical examples of non-convex optimization in ML are

Neural networks; maximum likelihood mixtures of Gaussians

- What is Gradient Descent Method?

- Tell us the difference between Batch Gradient Descent and Stochastic gradient descent.

- Give examples of some convex optimization problems in machine learning

- Give examples of the algorithms using Gradient based methods of second order information.

- Does Gradient Descent methods always converge to the same point?

- Is it necessary that the Gradient Descent Method will always find the global minima?

- What is a local optimum is and why is it important in a specific context, such as k-means clustering. What are specific ways for determining if you have a local optimum problem? What can be done to avoid local optima? Read possible answer

<u>Suggested Reading</u>
- Explain the Newton's method?

<u>Suggested Reading</u>
- What kind of problems are well suited for Newton's method? BFGS? SGD?

- What are "slack variables"?

- Describe a constrained optimization problem and how you would tackle it.

# Recommendation
some good examples of recommender models can be found <u>here</u>
- What is a recommendation engine? How does it work?

- How to do customer recommendation?

- What is Collaborative Filtering?

- How would you generate related searches for a search engine?

- How would you suggest followers on Twitter?

- Do you know about the Netflix Prize problem? How would you approach it?

Here is a nice post on the <u>Netflix challenge</u>

# Feature Engineering
Here is a <u>good article</u> on feature engineering
- What is Feature Engineering?

How predictors are encoded in a model can have a signi?cant impact on model performance and we achieve such encoding through feature engineering. Sometimes using combinations of predictors can be more e?ective than using the individual values: the product of two predictors may be more e?ective than using two independent predictors. Often the most e?ective encoding of the data is captured by the modeler's understanding of the problem and thus is not derived from any mathematical technique.
These features can be extracted in two ways: 1. By a human expert (known as hand-crafted) or 2. By using automated feature extraction methods such as PCA, or Deep Learning tools such as DBN. Both 1 and 2 can be used on top of each other.
- Give an example where feature example can be very useful in predicting results from data and explain with reason why it is so effective in some cases?

- What are some good ways for performing feature selection that do not involve exhaustive search?

7

- How to convert categorical variables to numerical for extracting features?

# Feature Selection

Here is a nice post on feature selection,
also known as variable selection, attribute selection or variable subset selection
- Explain feature selection and its importance with examples.

- What is variance threshold approach?

- How Univariate feature selection works?

- Is there any negative impact of using too many or too few variables?

- Is there any thumb rule for the number of features that should be used? How do you select the best features?

- What will be your approach to recursive feature elimination?

- Describe some feature selection methods.

- Does the model affect the choice of feature selection method?

# Natural Language Processing (NLP)

For basic introduction visit the wiki page.
Here is the link to coursera course for NLP
Pick the software from the The Stanford NLP (Natural Language Processing) Group and input some text to view its parse tree, named entities, part of speech tags, etc.
If the company deals with text data, you can expect some questions on NLP and Information Retrieval:
- Explain NLP to a non-technical person.

- What's the use of NLP in Machine Learning?

Some interesting usages are in areas like sentiment analysis, spam detecting, POS, Text summarization, Language translation etc.
- How unstructured text data can be converted into structured data for the purpose of ML models?

- Explain Vector Space Model and its use?

- Explain the distances and similarity measures that can be used to compare documents?

- Explain cosine similarity in a simple way? Suggested Reading

- Why and when stop words are removed? In which situation we do not remove them?

# Image processing and Text mining
- What tool would you prefer for image processing?

Some popular tools are: MATLAB, OpenCV or Octave

- What parameters would you consider while selecting a tool for image processing?

Ease of use, speed and resources needed are some of the common parameters
- How to apply Machine Learning to images?

- What are the text mining tools you are familiar with?

Some example are:
Commercial: Autonomy, Lexalytics , SAS/SPSS, SQLServer 2008+
OpenSource: RapidMiner , NClassifier, OpenTextSumarizer, WordNet, OpenNLP/ SharpNLP, Lucene/Lucene.NET, LingPipe, Weka
- What techniques do you apply for processing texts? Explain with an example.

- How to apply Machine Learning to audio data?

# Meta Learning
Wiki link on meta learning
- How will you differentiate between boosting and inductive transfer?

# Model selection
- What criteria would you use while selecting the best model from many different models?

- You have one model and want to find the best set of parameters for this model. How would you do that?

- How would you use model tuning for arriving at the best parameters? Suggested Reading

- Explain grid search and how you would use it?

- What is Cross-Validation?

- What is 10-Fold CV?

- What is the difference between holding out a validation set and doing 10-Fold CV?

# Evaluating Machine Learning
- How do you know if your model overfits?

- How do you assess the results of a logistic regression?

- Which evaluation metrics you know? Something apart from accuracy?

- Which is better: Too many false positives or too many false negatives?

- What precision and recall are?

- What is a ROC curve? Write pseudo-code to generate the data for such a curve.

- What is AU ROC (AUC)?

- Do you know about Concordance or Lift?

## Discussion Questions
- You have a marketing campaign and you want to send emails to users. You developed a model for predicting if a user will reply or not. How can you evaluate this model? Is there a chart you can use?

# Miscellaneous
## Curse of Dimensionality
- What is Curse of Dimensionality? What is the difference between density-sparse data and dimensionally-sparse data? Suggested Reading

- Dealing with correlated features in your data set, how to reduce the dimensionality of data.

- What are the problems of large feature space? How does it affect different models, e.g. OLS? What about computational complexity?

- What dimensionality reductions can be used for preprocessing the data?

- What is the difference between density-sparse data and dimensionally-sparse data?

- Others

- You are training an image classifier with limited data. What are some ways you can augment your dataset?

## 参考文章：
1. Collection of Machine Learning Interview Questions

## 资源推荐：
**1.优质在线课程 -《机器学习项目实战班》，网址：** http://www.jiuzhang.com/course/19/
学习 KNN, Naive Bayes, Logistic Regression, Neural Network, Deep Learning, Decision Tree, Bagging, Boosting 等机器学习算法；实战垃圾邮件分类、员工跳槽概率预测、商品类别判断等10余个实战项目。

## 2.优质公众号 - 机器学习与人工智能，公众号ID: machinelearningai

专注机器学习和人工智能，关注前沿技术和业界实践，旨在提供一线资源和消息。这里有最热门的新闻，这里有最专业的文章，这里有最具有价值的干货。