General Questions: Covering The Basics

These questions evaluate the basic understanding of machine learning in interviewees. These questions are usually relevant to candidates who are beginners and trying to get an entry-level position in data science. Here are some of the questions with answers that the candidates can prepare for:

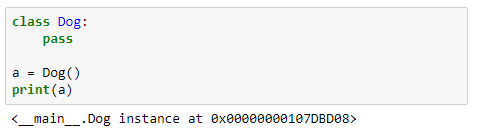
* *How is ML different from artificial intelligence?*
  + AI involves machines that execute tasks which are programmed and based on human intelligence, whereas ML is a subset application of AI where machines are made to learn information. They gradually perform tasks and can automatically build models from the learnings.
* *Differentiate between statistics and ML.*
  + In statistics, the relationships between relevant data (variables) is established; but in ML, the algorithms rely on data regardless of their statistical influence. In other words, statistics is concerned about inferences in the data whereas ML looks at optimisation.
* *What are neural networks and where do they find their application in ML? Elaborate.*
  + Neural networks are information processing models that derive their functions based on biological neurons found in the human brain. The reason they are the choice of technique in ML is because, they help discover patterns in data that are sometimes too complex to comprehend by humans.
* *Differentiate between a parameter and a hyperparameter?*
  + Parameters are attributes in training data that can be estimated during ML. Hyperparameters are attributes that cannot be determined beforehand in the training data. Example: Learning rate in neural networks.
* *What is ‘tuning’ in ML?*
  + Generally, the goal of ML is to automatically provide accurate output from the vast amounts of input data without human intervention. Tuning is a process which makes this possible and it involves optimising hyperparameters for an algorithm or a ML model to make them perform correctly.
* *What is optimisation in ML?*
  + Optimisation in general refers to minimising or maximising an objective function (in linear programming). In the context of ML, optimisation refers to tuning of hyperparameters which result in minimising the error function (or loss function).
* *What is the use of gradient descent?*
  + The use of gradient descent plainly lies with the fact that it is easy to implement and is compatible with most of the ML algorithms when it comes to optimisation. This technique works on the principle of [cost function](https://www.thedataschool.co.uk/benedetta-tagliaferri/machine-learning-101-cost-function/).
* *Explain any data preprocessing technique for ML.*
  + Standardisation: It is mainly used for algorithms following a Gaussian distribution. It can be done through scikit learn [Standardscaler](https://machinelearningmastery.com/prepare-data-machine-learning-python-scikit-learn/) class (for Python).
* *What is dimensionality reduction? Explain in detail.*
  + The process of reducing variables in a ML classification scenario is called Dimensionality reduction. The process is segregated into sub-processes called feature extraction and feature selection. Dimensionality reduction is done to enhance visualisation of training data. It finds the appropriate set of variables known as principal variables.
* *Explain Principal Component Analysis (PCA).*
  + PCA is a dimensionality-reduction technique which mathematically transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components.
* *What value do you optimise when using a support vector machine (SVM)?*
  + For a linear function, SVM optimises the product of input vectors as well as the coefficients. In other words, the algorithm with the linear function can be restructured into a dot-product.
* *On what basis do you choose a classifier?*
  + Classifiers must be chosen based on the accuracy it provides on the trained data. Also, the size of the dataset sometimes affects accuracy. For example, [Naive Bayes](https://www.youtube.com/watch?v=M59h7CFUwPU)classifiers suit smaller datasets in terms of accuracy due to higher [asymptotic errors](https://www.researchgate.net/post/What_is_the_difference_between_asymptotic_standard_error_and_standard_error).
* *Which is better for image classification? Supervised or unsupervised classification. Justify.*
  + In a supervised classification, the images are interpreted manually by the ML expert to create feature classes whereas this is not the case in unsupervised classification wherein the ML software creates feature classes based on image pixel values. Therefore, it is better to opt for supervised classification for image classification in terms of accuracy.
* *Mention key business metrics that help ML (company-specific).*
  + Identify the key services/products/functions that holds good for ML. For example, if you consider a commercial bank, metrics such as number of new accounts, type of accounts, leads generated and so on, can be evaluated through ML methods.

These questions are usually encountered when you face an ML job interview. It is also suggested to go through all topics related to ML since it spans a large number of concepts and techniques.

Programming Questions: Working Through The Code

ML also involves a significant amount of programming. The popular programming languages in this field are [Python](https://analyticsindiamag.com/can-learning-python-alone-lead-to-a-viable-data-science-career-path/) and R. It is important that you have a good command over these languages — preferably both. Here are a few questions that are generally asked:

**Python**

* *Is Python an object-oriented programming language? If yes, give an example.*
  + Yes.                                  [](https://analyticsindiamag.com/wp-content/uploads/2018/06/pythcode.png)
* *Mention the Python modules for numerical and scientific computations.*
  + [Numpy](https://analyticsindiamag.com/beginners-guide-numpy-must-python-library-data-scientist-toolkit/)
  + Pandas
  + Sci-Kit Learn
* *List the file processing modes supported in Python?*
  + Read-only mode (“r”)
  + Write-only mode (“w”)
  + Read-write mode (“rw”)
  + Append mode (“a”)
* *Which do you prefer? An integrated development environment (IDE) for Python or a notebook software such as Jupyter. Explain*
  + I prefer Jupyter Notebook over IDEs since its gives us the power to execute code step by step and to debugs the errors easily.
* *Name the commonly used libraries in Python for ML.*
  + Tensorflow
  + NLTK
  + Matplotlib
* *Differentiate tuples and lists in Python.*
  + Tuple is an immutable type which cannot be changed (it is static) whereas Lists are mutable (dynamic). Tuples do not support ‘append’ or ‘remove’ whereas Lists can.

**R**

* *List down the data structures supported in R.*
  + Atomic vector
  + List
  + Matrix
  + Data frame
  + Factors
  + Table
* *How do you create linear models in R?*
  + Using the lm() function
* *How do you test code written in R?*
  + Using a package called ‘[testthat](https://github.com/r-lib/testthat)’.
* *How will you read a .csv file in R?*
  + Using read.csv() function
* *What is the use of functions in R?*
  + Functions provide twofold advantage
    - Variable input can be used for different data
    - Output is returned as an object which enables the manipulation of the function.
* *What is ‘workspace’ in R?*
  + Workspace represents the interface that contains user-defined objects — vectors, lists, tables etc.

These questions represent a small part of a sea of questions available to ask in Python and R. The interviewee should be aware of all the relevant syntax related to ML and should be able to write codes comfortably.

Apart from these questions, the interviewee may be asked to write a pseudocode (either in Python or R) to solve a problem. It is suggested he or she practice programming to get a gist of its usage and syntax in ML.

Perceptive Questions: Grasping The Practicality In The Business

These questions make the interviewee think a deeper about his or her approach and see why there is a need for ML methods according to their perspective. This area tests the interviewee’s experience with ML. The following are a few questions asked considering the practical business aspect of ML.

* What does ML try to solve in the business?
* Does the ML model make any assumptions regarding the data for the project?
* Is the ML model quicker and efficient than conventional methods?

Questions like the ones above are mostly subjective in nature. The answer to these questions lies in the candidate’s ability and experience in using ML in the long run. There are no hard and fast rule answers to these questions. The interviewee should answer according to the context presented to them (usually type of company such as service or product-oriented businesses). Usually, these are asked for professionals with prior experience in ML.

Conclusion

Acing an ML job interview is a harrowing task worth accomplishing. The candidates will be tested thoroughly in their knowledge of ML since these future employees take care of deployments surrounding production and services related to ML. The questions mentioned above will give an idea of where to concentrate in a job interview related to ML and data science. Ultimately, the candidate should present his or her unique selling point apart from regular preparations for the interview.

All the best for your ML career!