Explain these things in your own words.

1) Let's say you have a set of data points y_i that are roughly linearly increasing as a functions of x_i. What does the linear regression model gives as a result in this case?

The model will give a set of values of y in the form of a straight line as x increases/decreases. Values of y are estimated as values of a function where x is the variable.

2) Explain in your own words what the Gradient Descent means.

It is an algorithm for optimizing Machine Learning models which involves an iteration of steps towards the negative direction of the differentiation of the cost function when training a model. In each iteration, the model parameters are tweaked little by little so that the cost function converges its minimum value. This cost function is defined by users, which can be anything but the most commonly used ones for linear regression tasks are Mean Squared Error, Mean Error and Mean Absolute Error.

3) How is the partial derivative related to the Gradient Descent?

The partial derivative represents the rate of change of a multivariable function with regard to each of the variable, in this case model parameters. Since the mechanism of gradient descent is basically to try with different combinations of model parameters to find one which produces the minimal cost function, it is ideal that when we change each of the parameters, we can see the movement direction of the function: if it is going down, it's good and this is what partial derivative is perfect for.

4) How is the Gradient Descent affected by different learning rates?

It is commonly the case that a smaller learning rate ensures a higher accuracy as the model changes slowly and thus does not skip the minimum value of the differential function. However, this also means it requires more time to learn as the learning rate is small. It is also worth keeping in mind that for data with complicated and non-linear pattern, the learning rate should also be large enough to escape the local minima.

5) Explain the Stochastic Gradient Descent method

This is a Gradient Descent method in which the gradients are iteratively calculated on a randomly chosen subset containing 1 instance of the dataset instead of the whole batch, which helps reducing the amount of temporary memory required to store the dataset when training models. However, since instances are gone through one by one, it is advised that they should be shuffled and distributed evenly so bias is not introduced to the trained models.

6) Explain polynomial regression

This is a variant of linear regression in which is the model to fit training data is of higher degree than 1 such as y=x^2 or y=x^3 (x is an independent variable, or i.e predictor and y is a dependent variable or a target). Since polynomial functions are more sensitive to change, polynomial regression algorithms perform better in practical cases where correlations between features are generally not linear, but it is also prone to overfitting if the model is too sensitive to small changes.

7) Explain underfitting and overfitting

Underfitting is a phenomenon in which the model failed to generalize well enough to have reliable predictions over new data. This might be because the model is not complex enough to model the data correlations or there are simply no correlations at all among the chosen features.

Overfitting is a phenomenon in which the model failed to generalize to perform well on a larger or new dataset. This could be because the model has not “learnt” enough so typically the solution is to train it with more data (if that is feasible), train with different train and validation set combinations before evaluate with test set or just use a simpler model.