Question Stochastic Gradient descent

Explain how Stochastic Gradient Descent works and what it aims for.

The main idea is finding a line that can model the given data.

The algorithm runs for a number of epochs visiting each observation of the data. For each observation the algorithm makes a prediction and this is compared with the actual class for the observation being processed to get an error before performing an update on the model.

- It is an optimisation algorithm to minimise a function by moving in the direction of the negative of the gradient of the function.
- It is used in ML for updating parameters or coefficients on a model such as regression or in logistic regression.
- In linear regression one possible strategy to find a solution is to minimize the mean squared error or MSE.

formula: $MSE = \frac{1}{N} = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$