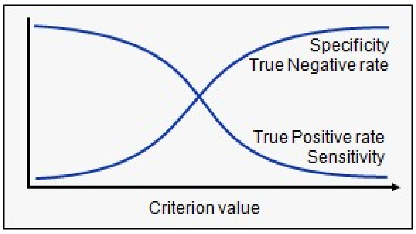
| | |
|---|---|
| **Regularization** | Regularization is a technique used to solve the overfitting problem in statistical models. In machine learning, regularization penalizes the coefficients such that the model generalize better. We have different types of regression techniques which uses regularization such as Ridge regression and lasso regression. |
| **Reinforcement Learning** | It is an example of machine learning where the machine is trained to take specific decisions based on the business requirement with the sole motto to maximize efficiency (performance). The idea involved in reinforcement learning is: The machine/ software agent trains itself on a continual basis based on the environment it is exposed to, and applies it's enriched knowledge to solve business problems. This continual learning process ensures less involvement of human expertise which in turn saves a lot of time!<br><br>Important Note: There is a subtle difference between Supervised Learning and Reinforcement Learning (RL). RL essentially involves learning by interacting with an environment. An RL agent learns from its past experience, rather from its continual trial and error learning process as against supervised learning where an external supervisor provides examples.<br><br>A good example to understand the difference is self driving cars. Self driving cars use Reinforcement learning to make decisions continuously like which route to take, what speed to drive on, are some of the questions which are decided after interacting with the environment. A simple manifestation for supervised learning would be to predict the total fare of a cab at the end of a journey. |
| **Residual** | Residual of a value is the difference between the observed value and the predicted value of the quantity of interest. Using the residual values, you can create residual plots which are useful for understanding the model. |
| **Response Variable** | Response variable (or dependent variable) is that variable whose variation depends on other variables. |
| **Ridge Regression** | Ridge regression performs '**L2 regularization**', i.e. it adds a factor of sum of squares of coefficients in the optimization objective. Thus, ridge regression optimizes the following:<br><br>Objective = RSS + α * (sum of square of coefficients)<br><br>Here, α (alpha) is the parameter which balances the amount of emphasis given to minimizing RSS vs minimizing sum of squares of coefficients. α can take various values:<br><br>1. **α = 0:**<br>○ The objective becomes same as simple linear regression.<br>○ We'll get the same coefficients as simple linear regression.<br>2. **α = ∞:**<br>○ The coefficients will be zero. This is because of infinite weightage on square of coefficients, anything less than zero will make the objective infinite.<br>3. **0 < α < ∞:**<br>○ The magnitude of α will decide the weightage given to different parts of objective.<br>○ The coefficients will be somewhere between 0 and 1 for simple linear regression. |

**ROC-AUC**

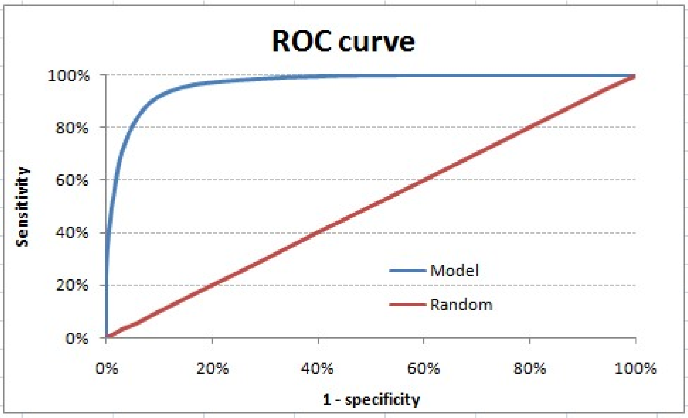
Let’s first understand what is ROC (Receiver operating characteristic) curve. If we look at the confusion matrix, we observe that for a probabilistic model, we get different value for each metric.

| Confusion Matrix | | Target | | | |
|---|---|---|---|---|---|
| | | Positive | Negative | | |
| **Model** | Positive | a | b | *Positive Predictive Value* | a/(a+b) |
| | Negative | c | d | *Negative Predictive Value* | d/(c+d) |
| | | *Sensitivity* | *Specificity* | **Accuracy** = (a+d)/(a+b+c+d) | |
| | | a/(a+c) | d/(b+d) | | |

Hence, for each sensitivity, we get a different specificity. The two vary as follows:




The ROC curve is the plot between sensitivity and (1- specificity). (1- specificity) is also known as false positive rate and sensitivity is also known as True Positive rate. Following is the ROC curve for the case in hand.




Let’s take an example of threshold = 0.5 (refer to confusion matrix). Here is the confusion matrix :

As you can see, the sensitivity at this threshold is 99.6% and the (1-specificity) is ~60%. This coordinate becomes on point in our ROC curve. To bring this curve down to a single number, we find the area under this curve (AUC).

Note that the area of entire square is 1*1 = 1. Hence, AUC itself is the ratio under the curve and the total area.

| | |
|---|---|
| **Root Mean Squared Error (RMSE)** | RMSE is a measure of the differences between values predicted by a model or an estimator and the values actually observed. It is the standard deviation of the residuals. Residuals are a measure of how far from the regression line data points are. The formula for RMSE is given by:<br>$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$<br>Here,<br>• Predicted -> value predicted by the model<br>• Actual -> observed values<br>• N -> Total number of observations |
| **Rotational Invariance** | In mathematics, a function defined on an inner product space is said to have rotational invariance if its value does not change when arbitrary rotations are applied to its argument. For example, the function:<br>$$f(x, y) = x^2 + y^2$$<br>is invariant under rotations of the plane around the origin, because for a rotated set of coordinates through any angle *θ*. |