# WORKSHEET

MACHINE LEARNING – WORKSHEET 3

**Q1 to Q15 are subjective answer type questions, Answer them briefly.**

1. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Answer: Linear regression identifies the equation that produces the smallest difference between all of the observed values and their [fitted values](https://statisticsbyjim.com/glossary/fitted-values/). To be precise, linear regression finds the smallest sum of squared [residuals](https://statisticsbyjim.com/glossary/residuals/) that is possible for the dataset

In machine learning, the **polynomial kernel** is a **kernel** function commonly used with **support vector machines** (**SVMs**) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over **polynomials** of the original variables, allowing learning of non-linear models.

In machine learning, the radial basis function **kernel**, or **RBF kernel**, is a popular **kernel** function

used in various kernelized learning algorithms. In particular, it is commonly used in support vector

machine classification.

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit of model in regression and why??

Answer: [R-squared](https://statisticsbyjim.com/glossary/r-squared/) is a goodness-of-fit measure for linear [regression](https://statisticsbyjim.com/glossary/regression-analysis/) models. This statistic indicates the percentage of the variance in the [dependent variable](https://statisticsbyjim.com/glossary/response-variables/) that the [independent variables](https://statisticsbyjim.com/glossary/predictor-variables/) explain collectively. R-squared measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 100% scale.

1. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

Answer: In statistical data analysis the **total sum of squares** (TSS or SST) is a quantity that appears as part of a standard way of presenting results of such analyses. For a set of observations , it is defined as the **sum** over all **squared** differences between the observations and their overall **mean** .

The **sum of squares** measures the deviation of data points away from the mean value. A higher **sum-of-squares** result indicates a large degree of variability within the data set, while a lower result indicates that the data does not vary considerably from the mean value.

1. R2 always takes on a value between 0 and 1. ...
2. is also known as the total sum of squares (TSS). ...
3. To **compute ESS**, you subtract the mean value of Y from each of the estimated values of Y; each term is squared and then added together:

A **residual sum of squares** (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model. ... The **residual sum of squares** measures the amount of error remaining between the regression function and the data set.

1. What is Gini –impurity index?

Answer: The **Gini impurity** measure is one of the methods used in decision tree algorithms to decide the optimal split from a root node, and subsequent splits. ... Def: **Gini Impurity** tells us what is the probability of misclassifying an observation. Note that the lower the **Gini** the better the split.

When training a decision tree, the best split is chosen by maximizing the **Gini** Gain, which is **calculated** by subtracting the weighted **impurities** of the branches from the original **impurity**.

1. Are unregularized decision-trees prone to overfitting? If yes, why?

Answer: **Decision trees** are **prone to overfitting**, especially when a **tree** is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

1. What is an ensemble technique in machine learning?

Answer: **Ensemble** methods **is** a **machine learning** technique that combines several base models in order to produce one optimal predictive model . ... A Decision Tree determines the predictive value based on series of questions and conditions.

1. What is the difference between Bagging and Boosting techniques?

Answer: **Bagging** is a way to decrease the variance **in the** prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. **Boosting** is an iterative technique which adjusts the weight of an observation based on the last classification.

If the problem is that the single model gets a very low performance, **Bagging** will rarely get a **better** bias. However, **Boosting** could generate a combined model with lower errors as it optimises the advantages and reduces pitfalls of the single model. ... For this reason, **Bagging** is effective more often **than Boosting**.

1. what is out-of-bag error in random forests?

Answer: **Out-of-bag** (**OOB**) **error**, also called **out-of-bag estimate**, is a method of measuring the prediction error of [random forests](https://en.wikipedia.org/wiki/Random_forest), [boosted decision trees](https://en.wikipedia.org/wiki/Gradient_boosting), and other [machine learning](https://en.wikipedia.org/wiki/Machine_learning) models utilizing [bootstrap aggregating](https://en.wikipedia.org/wiki/Bootstrap_aggregating) (bagging) to sub-sample data samples used for training. OOB is the mean prediction error on each training sample *xᵢ*, using only the trees that did not have *xᵢ* in their bootstrap sample.

1. What is K-fold cross-validation?

Answer: **Cross**-**validation** is a resampling procedure **used** to evaluate machine learning models on a limited data sample. The procedure has a single parameter called **k** that refers to the number of groups that a given data sample is to be split into.

1. What is hyper parameter tuning in machine learning and why it is done?

Answer: In **machine learning**, **hyperparameter optimization** or **tuning** is the problem of choosing a set of optimal **hyperparameters** for a **learning** algorithm. ... These measures are called **hyperparameters**, and have to be tuned so that the model can optimally solve the **machine learning** problem.

**Tuning** is the process of maximizing a model's performance without overfitting or creating too high of a variance. In **machine learning**, this is accomplished by selecting appropriate “hyperparameters.” Hyperparameters can be thought of as the “dials” or “knobs” of a **machine learning** model.

**Hyperparameters** are **important** because they directly control the behaviour of the training algorithm and have a **significant** impact on the performance of the model is being trained. “A good choice of **hyperparameters** can really make an algorithm shine”. ... Easy to manage a large set of experiments for **hyperparameter tuning**.

1. What issues can occur if we have a large learning rate in Gradient Descent?

Answer: The **learning rate** hyperparameter controls the **rate** or speed at which the model learns. ... A **learning rate** that is too small may never converge or may get stuck on a suboptimal solution. When the **learning rate** is too large, **gradient descent** can inadvertently increase rather than decrease the **training** error.

Deep **learning** neural networks are trained using the stochastic **gradient descent** algorithm. ... Specifically, the **learning rate** is a configurable hyperparameter used in the **training** of neural networks that has a small positive value, often in the range between 0.0 and 1.0.

1. What is bias-variance trade off in machine learning?

Answer: You now know that: **Bias** is the simplifying assumptions made by the model to make the target function easier to approximate. **Variance** is the amount that the estimate of the target function will change given different training data. **Trade-off** is tension between the error introduced by the **bias** and the **variance**.

1. What is the need of regularization in machine learning?

Answer: This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages **learning** a more complex or flexible model, so as to avoid the risk of overfitting. A simple relation for linear regression looks like this.

**Regularization** is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values.

1. Differentiate between Adaboost and Gradient Boosting?

Answer: **Adaboost** is more about 'voting weights' and **gradient boosting** is more about 'adding **gradient** optimization'. **Adaboost** doesn't overfit because it is more about 'organizing people to vote' than 'voting'. In fact, if you have a **gradient boosting** model, you can use it in **adaboost** along with other models.

1. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Answer: **Logistic regression** is known and **used** as a **linear classifier**. ... It is **used** to come up with a hyperplane in feature space to separate observations that belong to a class from all the other observations that **do** not belong to that class. The decision boundary is thus **linear**.

