# Regularization

* Used to avoid overfitting problem, by explicitly controlling the model complexity and changing the loss function. by shrinking (imposing a penalty) on the coefficients.
* L1- LASSO Regression, L2-Ridge Regression
* The main difference between L1 and L2 regularization is amount of penalty added to the coefficients.

**Difference between L1 and L2**  
- L2 (Ridge) shrinks all the coefficient by the same proportions but eliminates none, while L1 (Lasso) can shrink some coefficients to zero, performing variable selection.

* L1- minimizes sum of absolute differences between dependent and independent variables.
* L2- minimizes sum of squared differences, between dependent and independent variables

**Which to use?**  
If all the features are correlated with the label, ridge (L2) outperforms lasso (L1), as the coefficients are never zero in ridge. If there’s lot of sparsity in the data, lasso outperforms ridge as in lasso model some coefficient can be shrunken to zero.

Gradient Boosting:

Machine learning techniques for regression and classification problem, which produces prediction model in the form of an ensemble of weak prediction models. It produces multiple prediction models by changing parameters.

Model is executed ‘n’ number of times (called bootstrapping) to minimize the error, either it uses local optima or local minima.

In every iteration tries to reduce the cost function and error.

PCA: principle component analysis

Is a dimensionality reduction technique. PCA yields the directions (principal components) that maximize the variance of the data. It eliminates the variables which contribute to high variance.

### **A Summary of the PCA Approach**

* Standardize the data.
* Obtain the Eigenvectors and Eigenvalues from the covariance matrix or correlation matrix, or perform Singular Vector Decomposition.
* Sort eigenvalues in descending order and choose the k eigenvectors that correspond to the k-largest eigenvalues where k is the number of dimensions of the new feature subspace (k≤dk≤d).
* Construct the projection matrix WW from the selected k eigenvectors.
* Transform the original dataset XX via WW to obtain a k-dimensional feature subspace YY.

***Note:***

*Singular value decomposition, LDA Linear Discriminant Analysis are also other dimensionality reduction techniques.*