Statistical Learning

- It refers to a set of tools for understanding data and these tools can be classified as supervised or unsupervised.
- These supervised techniques are used to build a statistical model to predict output based on the input.
- And Unsupervised techniques are used to learn relationships and structure from such data.

@learn.machinelearning

History of Statistical Learning

- At the beginning of the nineteenth century, Legendre and Gauss published papers on the method of least squares, which implemented the earliest form of what is now known as linear regression.

- Fisher proposed linear discriminant analysis in 1936. In the1940s, various authors put forth an alternative approach, logistic regression.
- In the early 1970s, Nelder and Wedderburn coined the term generalized linear models for an entire class of statistical learning methods that include both linear and logistic regression as special cases.

History of Statistical Learning

- Mostly in 1970s because of computation power linear models are famous than non linear models.
- By the 1980s, computing technology had finally improved sufficiently that non-linear methods were no longer computationally prohibitive.

- 1980s Breiman, Friedman, Olshen and Stone introduced classification and regression trees, and a detailed practical implementation of a method, including cross-validation for model selection.
- Hastie and Tibshirani coined the term generalized additive models in 1986 for a class of non-linear extensions to generalized linear models

What Is Statistical Learning

- Lets consider a supervised problem where we know the input(X) and output(Y).
- We want to find the relationship between input(X) and output(Y), which can be written as $Y = f(X) + e$ where e is some random error.
- In this formula- error termtion, f represents the systematic information that X provides about Y.

- However, the function f that connects the input variable to the output variable is in general unknown. In this situation one must estimate f based on the observed points.
- In essence, statistical learning refers to a set of approaches for estimating f.

What Is Statistical Learning

- There are two main reasons that we may wish to estimate f: prediction and inference.
- prediction: In many situations we just know the input variable and unknow output variable. Here we can use Y = f(X) to obtain Y.
- Here f is treated as black box that means, one is not typically concerned with the exact form of f.

- Were f is not always the perfect function which maps the relation between input and output.
- And the accuracy of Y will be depending on two things reducible error and irreducible error.

What Is Statistical Learning

- In general, f will not be a perfect estimation, and this inaccuracy will introduce some error.
- This error is reducible because we can potentially improve the accuracy of f by using the most appropriate statistical learning technique to estimate f.
- Even if it were possible to form a perfect estimate for f, our prediction would still have some error in it! This is because Y is also a function of e, which, by definition, cannot be predicted using X.
- Therefore, variability associated with e also affects the accuracy of our predictions. This is known as the irreducible error, because no matter how well we estimate f, we cannot reduce the error introduced by e.

What Is Statistical Learning

- So why we have irreducible error?
- The quantity e may contain unmeasured variables that are useful in predicting Y. since we don't measure them, f cannot use them for its prediction.
- The quantity e may also contain unmeasurable variation. Which means it will be always different for given data(If given data changes).
- That irreducible error is always unknown.

What Is Statistical Learning

- **Inference:** We want to understand the relationship between X and Y , or more specifically, to understand how Y changes as a function of X.
- Now f cannot be treated as a black box, because we need to know its exact form.
- By doing this we can answer questions like (Which predictors are associated with the response?, What is the relationship between the response and each predictor?, is the relation is linear or complicated?)

Statistical learning

- **Parametric Methods:** It involve a two-step model-based approach.
- First, we make an assumption about the functional form, or shape, of f.
- For example, one very simple assumption is that f is linear in X:
- $f(X) = \beta 0 + \beta 1 X 1 + \beta 2 X 2 + \ldots + \beta p X p.$
- Instead of having to estimate an entirely arbitrary n-dimensional function f(X), one only needs to estimate the n + 1 coefficients $\beta 0, \beta 1, \ldots, \beta p$.

Statistical learning

- Second, After selecting a model to use we need a procedure that uses the training data to fit or train the model. In the case of the linear model, we need to estimate the parameters $\beta 0$, $\beta 1$,...,βp. That is, we want to find values of these parameters such that.
- $$Y \approx \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta pXp.$$
- There are may ways to estimate these parameters. The model-based approach just described is referred to as parametric.
- It reduces the problem of estimating f down to one of estimating a set of parameters

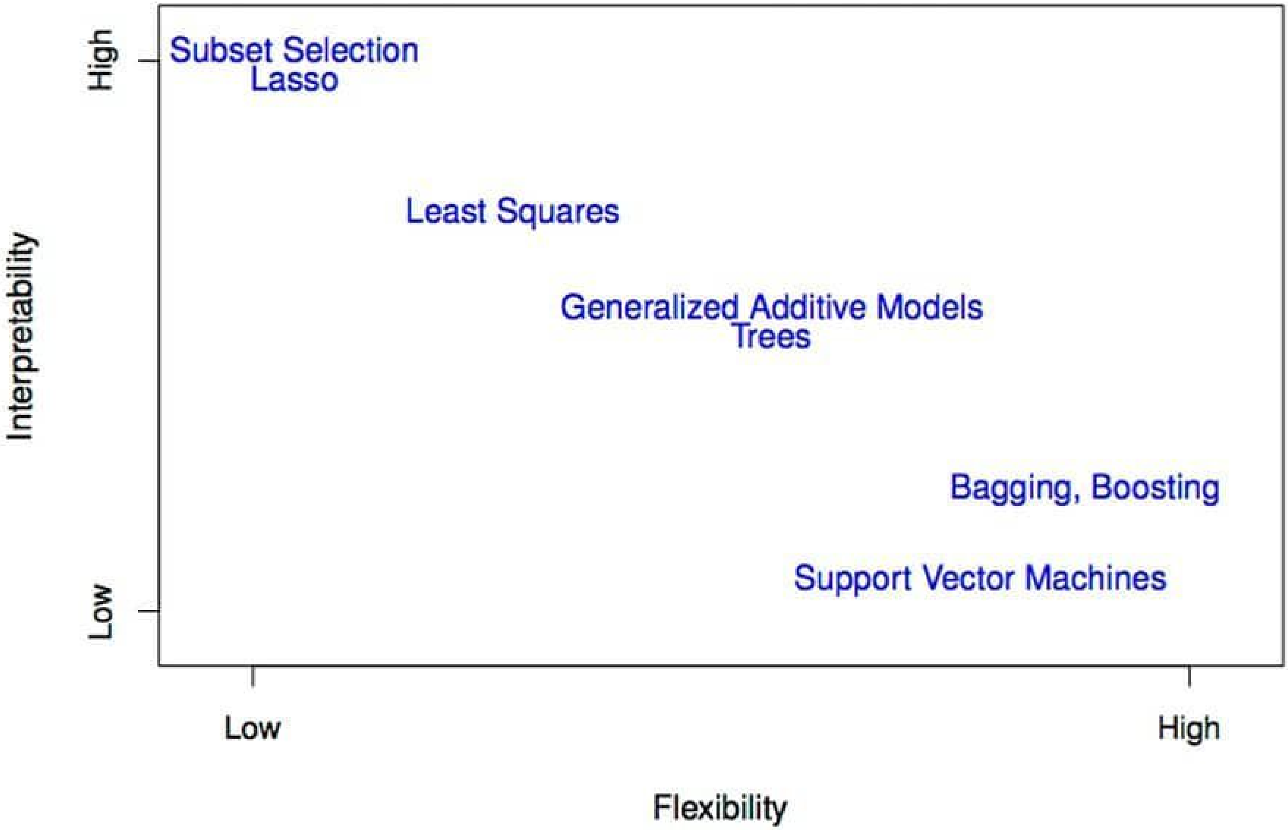
Statistical learning

- Assuming a parametric form for f simplifies the problem of estimating f because it is generally much easier to estimate a set of parameters, such as $\beta 0$, $\beta 1,\ldots,\beta p$ in the linear model, than it is to fit an entirely arbitrary function f.
- The potential disadvantage of a parametric approach is that the model we choose will usually not match the true unknown form of f.
- If the chosen model is too far from the true f, then our estimate will be poor.
- We can try to address this problem by choosing flexible models that can fit many different possible functional forms for f.

Statistical learning

- But in general, fitting a more flexible model requires estimating a greater number of parameters.
- These more complex models can lead to a phenomenon known as overfitting the data, which essentially means they follow the errors, or noise, too closely.

Trade-Off Between Prediction Accuracy and Model Interpretability



- **When interpretability is the goal, there are clear advantages to using simple and relatively inflexible statistical learning methods and vice versa.**

Statistical learning

- Non-parametric Methods: It do not make explicit assumptions about the functional form of f.
- Instead they seek an estimate of f that gets as close to the data points as possible without being too rough or wiggly.
- Such approaches can have a major advantage over parametric approaches: by avoiding the assumption of a particular functional form for f, they have the potential to accurately fit a wider range of possible shapes for f.

Statistical learning

- Any parametric approach brings with it the possibility that the functional form used to estimate f is very different from the true f, in which case the resulting model will not fit the data well.

- But non-parametric approaches completely avoid this danger, since essentially no assumption about the form of f is made.
- But non-parametric approaches do suffer from a major disadvantage: since they do not reduce the problem of estimating f to a small number of parameters, a very large number of observations is required in order to obtain an accurate estimate for f.