ACTL3142 Week 2 - Simple Linear Regression

Tadhg Xu-Glassop

2025T2

Hello!

My name is Tadhg (tie - gh) and I'm excited to be tutoring you this term!

All resources used in tutorials will be uploaded to my github `@txuglassop`.

Advice: This course is far too rich to be learnt exclusively within tutorials. Please keep up with lectures and preferably read the textbook. The exam is also typed, so do not focus on technical/mathematical aspects, but rather intuition and application.

Recap from Week 1

Hopefully you know what:

- statistical/machine learning is;

Recap from Week 1

Hopefully you know what:

- statistical/machine learning is;
    - Relate an input matrix (dataset) X to a response variable Y with $Y = f(X) + \epsilon$, where we predict f. Resultingly, $\hat{f}$ is said to be our model.
- supervised / unsupervised learning is;

Recap from Week 1

Hopefully you know what:

- statistical/machine learning is;
  - Relate an input matrix (dataset) X to a response variable Y with $Y = f(X) + \epsilon$, where we predict f. Resultingly, $\hat{f}$ is said to be our model.
- supervised / unsupervised learning is;
  - supervised has a response variable, whereas unsupervised does not.
- noise, overfit/underfit, training/test set, regression/classification problems, predictors/response are.

Recap from Week 1

Hopefully you know what:

- statistical/machine learning is;
  - Relate an input matrix (dataset) X to a response variable Y with $Y = f(X) + \epsilon$, where we predict f. Resultingly, $\hat{f}$ is said to be our model.
- supervised / unsupervised learning is;
  - supervised has a response variable, whereas unsupervised does not.
- noise, overfit/underfit, training/test set, regression/classification problems, predictors/response are.
- model interpretability and flexibility are.

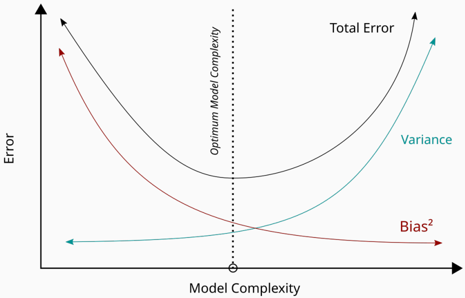
Recap from Week 1

Hopefully you know what:

- statistical/machine learning is;
  - Relate an input matrix (dataset) X to a response variable Y with $Y = f(X) + \epsilon$, where we predict f. Resultingly, $\hat{f}$ is said to be our model.
- supervised / unsupervised learning is;
  - supervised has a response variable, whereas unsupervised does not.
- noise, overfit/underfit, training/test set, regression/classification problems, predictors/response are.
- model interpretability and flexibility are.
  - interpretability of a model is how easy it is to make **inferences** about the nature of f and relationships between X and Y. Flexibility is the capability of the model to capture complex relationships.

Bias-Variance Tradeoff

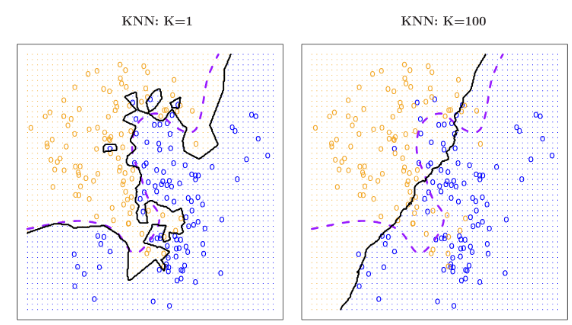
It can be shown that the MSE of a model is the sum of (squared) bias, variance and irreducible error.



More flexible $\implies$ More variance, less bias.
Less flexible $\implies$ More bias, less variance.

The best model for a problem achieves an optimal tradeoff between the two.

BV-Tradeoff Example - KNN



$K = 1 \Rightarrow$ More flexible $\Rightarrow$ Low bias, high variance $\Rightarrow$ overfit
$K = 100 \Rightarrow$ Less flexible $\Rightarrow$ High bias, low variance $\Rightarrow$ underfit

There is some K between 1 and 100 that achieves a better bias-variance tradeoff!

(Simple) Linear Regression

Simple Linear Regression

Relate a *single* covariate to the response with a linear function. That is,

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \epsilon.$$

Can you interpret the β coefficients?

(Simple) Linear Regression

Simple Linear Regression

Relate a *single* covariate to the response with a linear function. That is,

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \epsilon.$$

Can you interpret the β coefficients?

- β_0 is the predicted value of y when x is 0.
- β_1 is the average change in y for a unit increase in x.

(Simple) Linear Regression

Simple Linear Regression

Relate a *single* covariate to the response with a linear function. That is,

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \epsilon.$$

Can you interpret the β coefficients?

- β_0 is the predicted value of y when x is 0.
- β_1 is the average change in y for a unit increase in x.

This model is NOT very flexible (we are imposing a linearity assumption), but it's easy to interpret!

Estimating Coefficients

Given a training set, we need assumptions for the ϵ term to estimate the β params.

Ordinary Least Squares (OLS) (weak assumptions)

- Assume errors have zero mean, constant variance and conditionally uncorrelated.
- Don't need distributional assumption for OLS.

MLE (strong assumptions)

- Assume $\epsilon_i \mid \mathbf{x} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.
- Need a distributional assumption to construct our likelihood function and later construct confidence intervals.

Assessing Model

Under strong assumptions (MLE), β's can follow a t-distribution, allowing us to make confidence intervals and perform hypothesis tests. Specifically,

$$\frac{\hat{\beta}_i - \beta_i}{\mathrm{S.E.}(\hat{\beta}_\mathrm{i})} \sim t_{n-2}.$$

Most of the time, we test $\mathcal{H}_0 : \beta_1 = 0$.

The R^2 stat is "the proportion of variance explained" by our model.