1. Definition:
   1. live data,
   2. Feature engineering
   3. Prediction error
   4. homoskedasticity
   5. standard error
   6. Standard Prediction Error
   7. Loss Function
   8. Squared loss
   9. MSE, MAE
   10. Prediction error
   11. Cross-Validation
   12. Stationary
   13. Predictive analytics, Machine learning, algorithm and formula.
2. How is prediction error decomposed?
3. How to decompose the prediction error? Which will be quantified by interval predictions?
4. What’s the relation between confidence level and prediction level?
5. Explain the symmetry and convexity of loss function.
6. How to decompose MSE? Is OLS biased?
7. What’s the criterion of model selection?
8. What’s overfitting? What’s the reason?
9. As an indirect way of choosing models of fit, can we use R-square or MSE? What’s the better way?
10. As a direct way of choosing models of fit (i.e., training and test samples), what should we use?
11. What’s the difference between adjusted r-square and r-square?
12. What are the advantages of BIC and RMSE?
13. Definition:
    1. Live data: Actual data that are employed during the final testing of a computer system, as opposed to test data.
    2. variable selection
    3. predicted value – actual value
    4. Homoskedasticity occurs when the variance of the error term in a regression model is constant. It is an important assumption of classic regression. The opposite is heteroskedasticity, usually in the form of e related to x.
    5. 标准误（SE）的“小点点”是很多次抽样得到的很多「样本量均为N」的样本（样本的某种统计量，如平均值、回归系数等）
    6. SE of prediction,

A picture containing text, clock

Description automatically generated

* 1. a method of evaluating how well your algorithm models your dataset. Better algorithm outputs higher number.
  2. The most widely used loss function, symmetric and convex.

L( ej ) = ej2 = ( yjˆ − yj )2

* 1. Mean Squared Error: A picture containing text, clock

     Description automatically generatedalso, the numerator of the R-squared

RMSE = square root of MSE, convex and symmetric.

* 1. y-hat minus average y
  2. 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. As such, the procedure is often called k-fold cross-validation.
  3. Stationarity means that the way variables are distributed remains the same over time.
  4. Predictive analytics is often used for data analysis whose goal is prediction. But a more popular, and related, term is machine learning.

Machine learning is an umbrella concept for methods that use algorithms to find patterns in data and use them for prediction purposes.

An algorithm is a set of rules and steps that defines how to generate an output (predicted values) using various inputs (variables, observations in the original data).

A formula is an example of an algorithm – one that can be formulated in terms of an equation.

1. direction of miss and size
2. three parts:
   1. estimation error
   2. model error
   3. genuine error (idiosyncratic or irreducible error, with best model and zero estimation error we still can’t estimate all predicted values)
   4. estimation and genuine error can be quantified by interval predictions.
   5. The CI of the predicted value is about y-hat: where to expect the average value of the dependent variable if we know xj;
   6. The PI (prediction interval) is about yj itself: where to expect the actual value of yj if we know xj.
   7. PI starts with CI. But adds additional uncertainty (Std[εi]) that actual yj will be around its conditional.
3. Chart, line chart

   Description automatically generated
   1. Symmetry: If losses due to errors in opposing direction are similar
   2. Convexity: If twice as large errors generate more than twice as large losses
4. decompose MSE into Bias + Variance
   1. The bias of a prediction is the average of its prediction error.
   2. The variance of a prediction describes how it varies around its average value when multiple predictions are made.
   3. Text, letter

      Description automatically generated
5. Model selection is finding the best fit while avoiding overfitting and aiming for high external validity.
6. Model 1 may actually give a better fit in the original, but a worse fit in the live data. In this case, we say that model 1 overfits the original data. The typical reason for overfitting is fitting a model:
   1. with too many predictor variables
   2. Specifying too many interactions
   3. Too detailed nonlinear patterns, such as piecewise linear splines with many knots, or polynomials of high degree
7. Don’t, use BIC or AIC instead.
8. Pick model with lowest avg RMSE
9. The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. In a word, the adjusted R-squared adds penalty to increased number of predictors.
10. The advantage of BIC is that it needs no sample splitting which may be a problem in small samples. The advantage of test RMSE is that it makes no assumption. We do BIC first and then do a test MSE.