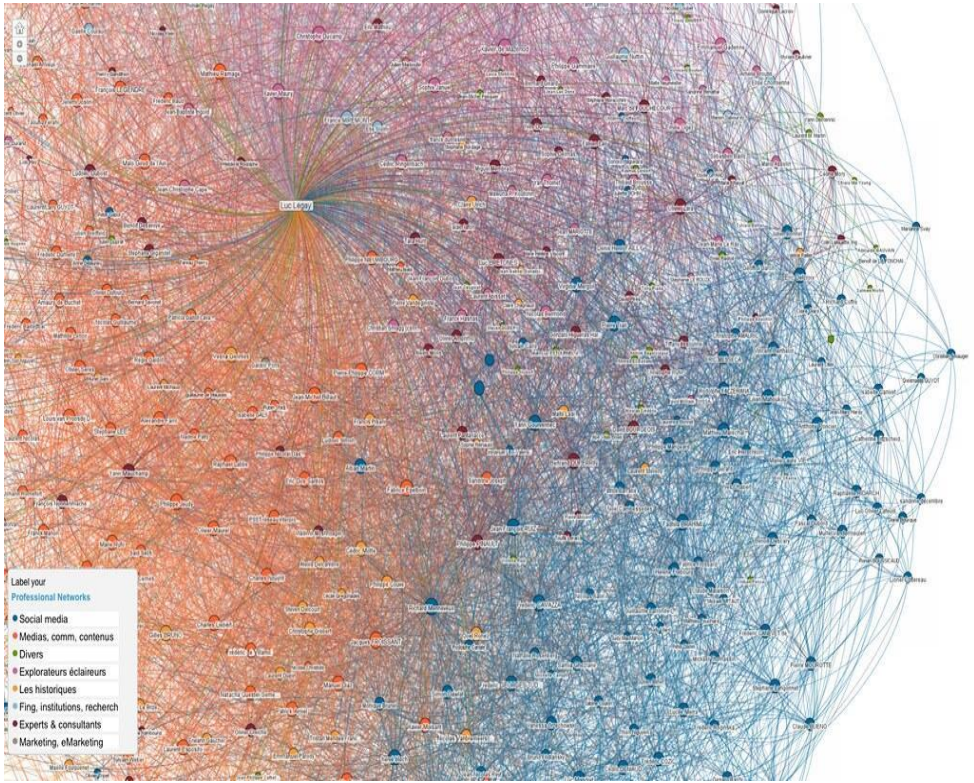
Ensemble Methods

Making Decision Trees More Powerful

Ensemble Methods

- Ways of combining different models together
- Allow you to take models with individual weakness and smooth them out
- Technique works very well with decision trees due to their erratic nature
- Highest performing ensemble method:
    - Boosting



Discussion: Ensemble Methods

You can use any base learner with an ensemble method, but trees are almost always the best choice.

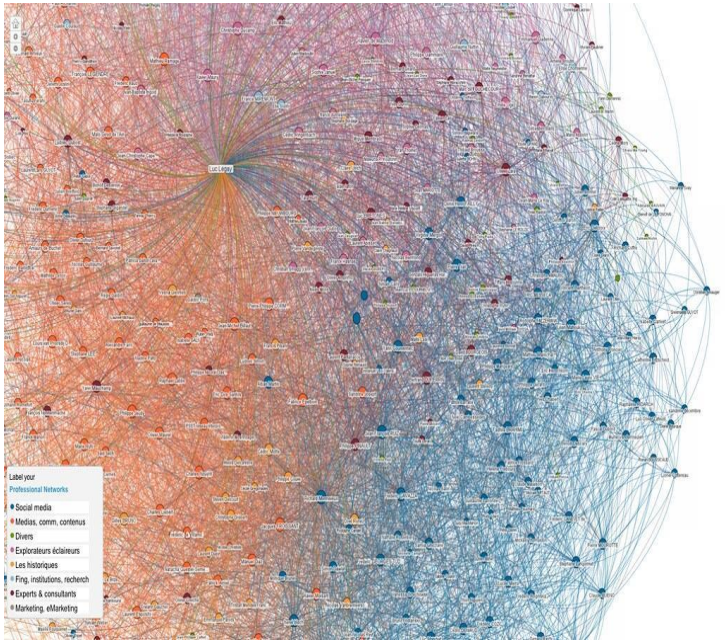
Why might that be the case?

Gradient Boosting

Making Decision Trees More Powerful

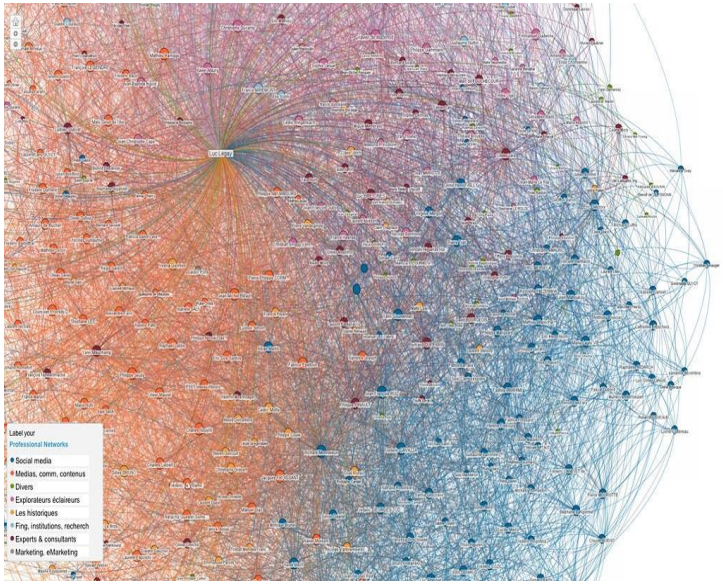
Gradient Boosting

- Boosting:
  - A way of combining many weak models into something more powerful
  - Is **additive:**
    - Means you add up the versions of different models together
  - Is **adaptive:**
    - Every subsequent model pays more attention to the answers that it got wrong on the last model



Gradient Boosting

- Gradients:
  - A fancy word for “errors”
- Ie, in gradient boosting, you don’t predict the answers for your model, but rather the amount of error it’s going to have, and then you add these values up
- Based around a very simple idea:
  - Correct answer = our guess + error
- In gradient boosting, you start with a naive guess, and continually predict the error from the previous model and combine the new error estimate with the previous model



Gradient Boosting:

- **Has established itself as the premiere out-of-the-box learning technique for structured data**
- **Gradient Boosted trees give you the best of many worlds:**
  - Non-linear, non-assumptive, highly accurate on many problems

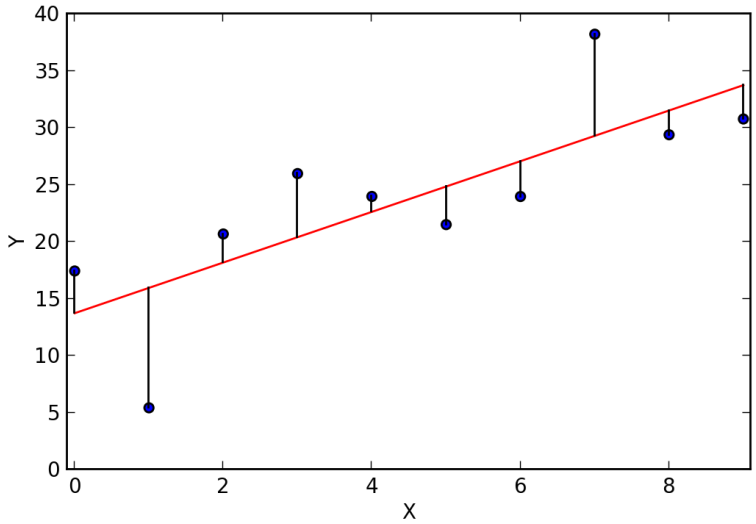
Model Interpretation

Machine Learning With Python

Model Interpretation - Common Metrics

R^2 Value

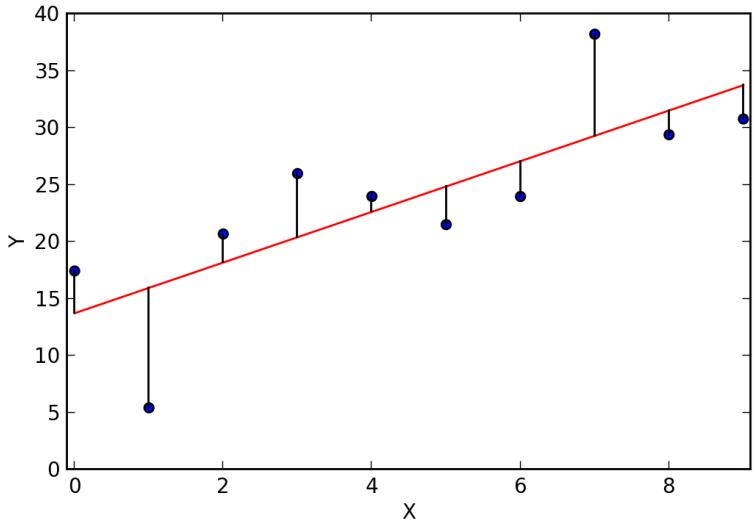
- How much of the change in y is described by the change in X
- Captures how much better your model is than the simplest alternative -- predicting the average



Model Interpretation - Common Metrics

Feature Importances

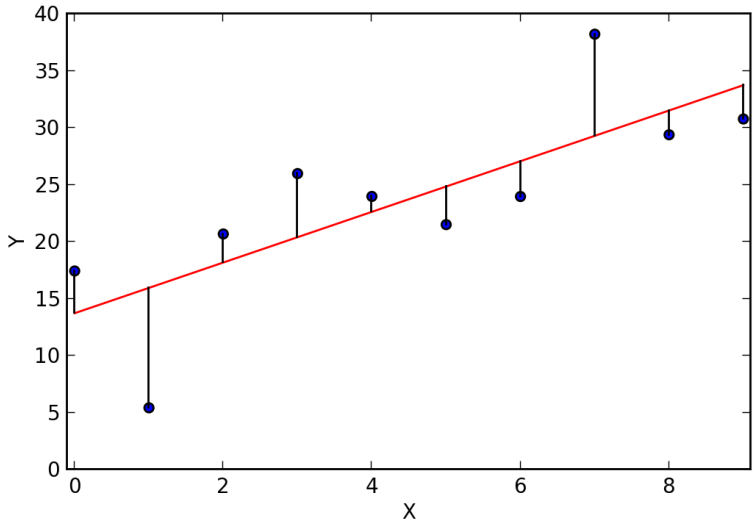
- Tells you how much each column contributed to predictive accuracy
- Demonstrates rank order importance of each column for your model



Model Interpretation - Common Metrics

Learning Rate

- Determines how quickly you update your model after each round of boosting
- One of the most important parameters for training a GBM



Discussion: Ensemble Methods

With a gradient boosted machine using trees as base learners, what are some parameters of our model that might affect how it performs?

Using intuition, what combination of attributes might we want to look for to give us the best results?

Discussion: Ensemble Methods

You can use any base learner with an ensemble method, but trees are almost always the best choice.

Why might that be the case?

