ETC3250

Business Analytics

Week 1.
Introduction to Business Analytics & R

29 July 2015

Example 1: Prediction

Kaggle competition: bulldozer auction prices

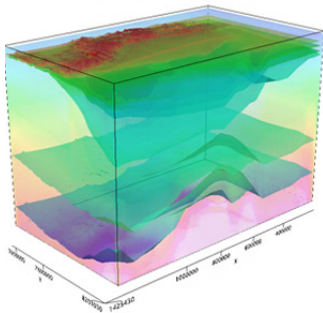**Problem**: Predict a bulldozer sale price based on known characteristics:

- Year made
- Sale date
- Machine hours
- Product group
- Type of enclosure

1. How could you improve the model?
2. How to use model to predict bulldozer prices when some information missing?
3. Does predicting log price make sense?

Example 2: Classification

Identification of high gold potentials

Problem: Identify high gold potentials at deep layers from soil properties measurements at multiple locations.

Example 2: Classification

```
Longitude Latitude    Ag   As   Ce   Co     Sr Ta Tb Te   Th     Mo    Au
   43.1723  22.3255 5.00  2.0 3.67 1.39  0.46  1  4  1 5.5    8.8  2.35
   43.1732  22.3255 0.63 15.5 1.33 3.94 12.95  1  1  2 1.5  160.0  4.29
   43.1742  22.3255 3.81  2.5 0.67 1.44  2.73  1  1  1 0.5    2.0 21.94
```

ADD Gold/No gold
ADD Low, Medium and High gold

1. How to apply multi-class classification with binary classifiers?
2. Can you give an example of unbalanced classification problem?
3. Does predicting log price make sense?