



Ensemble methods

Jan 23rd 2020

Ensemble Methods

- Bagging
- Boosting
- Stacking



Choosing threshold

Task: Work with your neighbor to choose the best threshold:

- **Threshold 1:** $TPR = 30\%$ $FPR = 5\%$
- **Threshold 2:** $TPR = 50\%$ $FPR = 20\%$
- **Threshold 3:** $TPR = 70\%$ $FPR = 40\%$

- Take into account the following prevalence and costs:
 - Prevalence: 60%
 - $Cost(FP) - Cost(TN) = 8$
 - $Cost(FN) - Cost(TP) = 4$

We will then come back to the large group and I'll pick some of you to share your answers with the rest of the class.



Bagging

- Bootstrap AGGREGatING
- Sampling rows with replacement
- Train a model on each of pulled datasets
- Each row counts the same (although some are repeated)
- Highly parallelizable
- Like checking with different experts (or doing a reference check)



Bagging

- Prediction
 - make each classifier generate a prediction (like a vote)
- Predict_proba
 - Mean of the predict_proba of the estimators inside it
- Relies on weak learners
 - The more different the better
- Doesn't improve performance
- Reduces chance of overfitting
- Example: Random Forest
 - (which samples variables too)

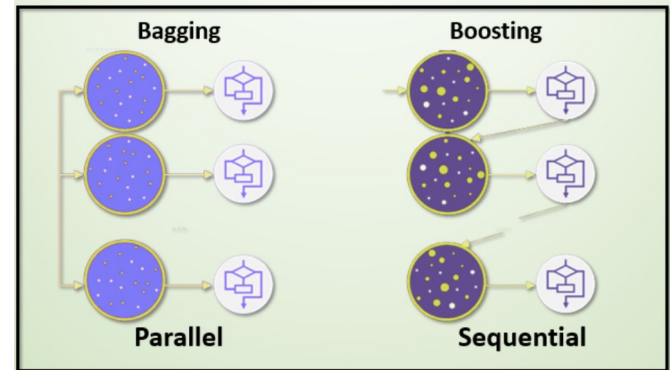


Boosting

- Take a sample of your data
- Train a learner
- Calculate your error
- Increase the likelihood of picking the rows that led to error
- Sample again (with your updated probabilities)
- Train another instance of the same learner
- Rinse and repeat

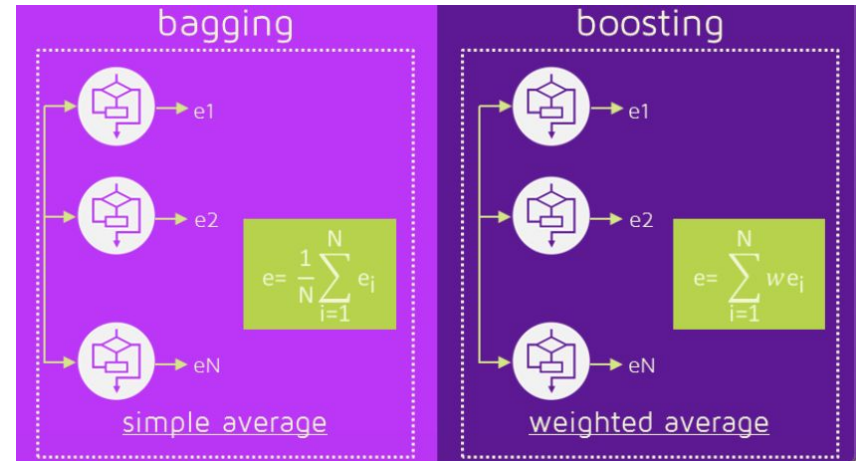


Bagging and Boosting



Boosting

- The model focuses on those cases that is finding most difficult
 - Improves performance
 - Just like a gritty student
- Sequential
 - You need to wait until you finish training a model to train the following
- Prediction: weighted average
- Examples:
 - Adaboost
 - XGBoost



Stacking

- Trains a learner that tries to predict which previously trained model works best for each datapoint
- It's common to see in use with multiple types of learners
- Prediction:
 - It matches each datapoint with the best model you have
 - Only one model for each datapoint
- Like a recruiter



Ensemble Methods Recap

Task: Discuss with your partner:

- How is sampling different in Bagging vs Boosting?
- How are the effects on performance different in Bagging vs Boosting?
- How are predictions generated differently in Bagging vs Boosting?
- What makes Stacking different?

We will then come back to the large group and I'll pick some of you to share your answers with the rest of the class.



Delivering Value

Your job is not to create high performance models

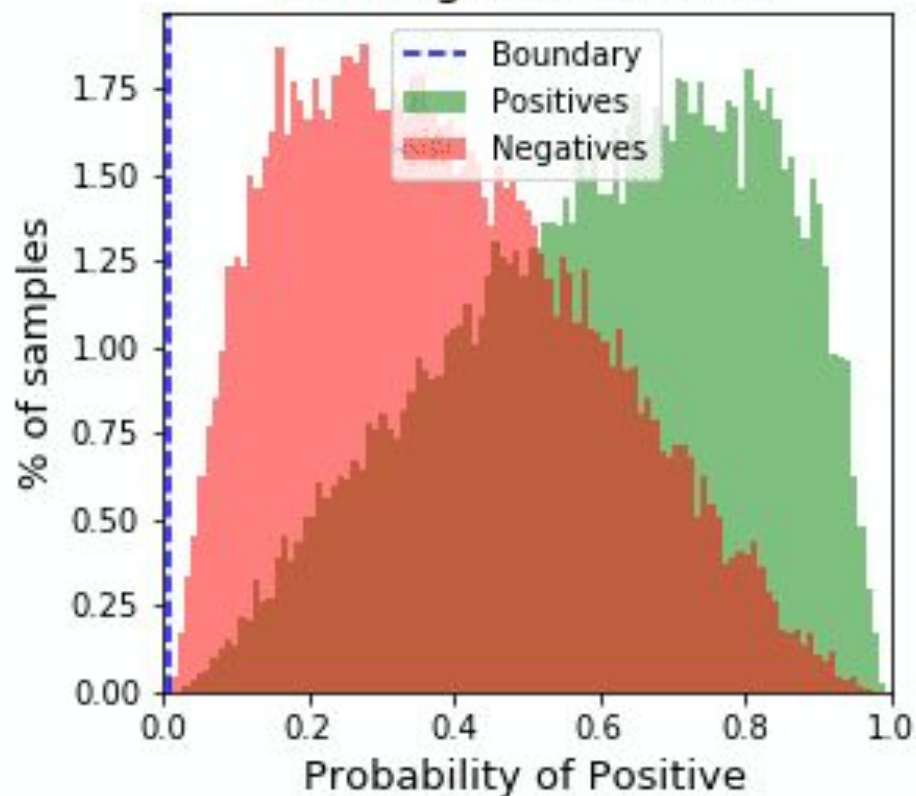
They pay you to **solve problems**

Summary + Exit Ticket

Presented by Dan Sanz

Thresholds

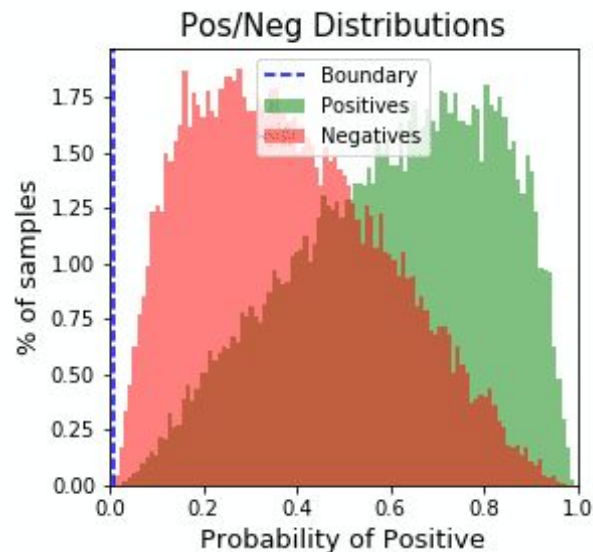
Pos/Neg Distributions



Confusion Matrix @threshold=0.01
power: 1.00 alpha: 1.00

Predicted Values	True Values	
	1	0
1	25063 True Positives	24937 False Positives
0	0 False Negatives	0 True Negatives

Thresholds to ROC curve



Confusion Matrix @threshold=0.01
power: 1.00 alpha: 1.00

