

Predicting Survival on the Titanic Using Gradient Boosted Trees



Swetha Kolalapudi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Understanding Gradient Boosted Trees

Use GBT to solve the Titanic problem

Understand the different parameters which can be used to control the ensemble

Find the best combination of parameters using hyper-parameter tuning

Gradient Boosted Trees

Models built using different



**Training
Sets**

**Each tree built
from a different
subset of the
training set**

Gradient Boosted Trees



Boosting

**Each tree built
sequentially from
a different subset
of the training set**



**Gradient
Descent**

**Adjusting
probability of a
data point being
in the next training
set**

Gradient Boosted Trees

Boosting

**Each tree built
sequentially from
a different subset
of the training set**

**Gradient
Descent**

Adjusting
probability of a
data point being
in the next training
set

Boosting

**In Random Forests,
each tree is built
independently**

Boosting

**With boosting, each
tree is built to
improve upon the
previous tree**

Boosting

Initially

**All data points have
equal probability of
being chosen**

Boosting

In each iteration

**The probabilities are
modified**

Boosting

**Data points which are
classified correctly
are down weighted**

**Data points which are
classified incorrectly
are up weighted**

Boosting

**In the final result,
each tree's vote is
given a different
weight**

Boosting

**Tree weight depends
on how accurate the
tree was on training
data**

Gradient Boosted Trees

Boosting

**Each tree built
sequentially from
a different subset
of the training set**

**Gradient
Descent**

Adjusting
probability of a
data point being
in the next training
set

Gradient Boosted Trees

Boosting

Each tree built
sequentially from
a different subset
of the training set

Gradient
Descent

Adjusting
probability of a
data point being
in the next training
set

Gradient
Descent

**The probabilities of
choosing a training
data point are
updated using this
technique**

Gradient Boosted Trees

A solid blue square.

Boosting

**Each tree built
sequentially from
a different subset
of the training set**

A solid orange square.

**Gradient
Descent**

**Adjusting
probability of a
data point being
in the next training
set**

Demo

**Use Gradient Boosted Trees to solve
the Titanic problem**

Demo

Find the best parameters using hyper-parameter tuning

Summary

Understanding Gradient Boosted Trees

Use GBT to solve the Titanic problem

**Understand the different parameters
which can be used to control the
ensemble**

**Find the best combination of
parameters using hyper-parameter
tuning**