

Introduction to MOOCs database

CS-411 : Digital Education & Learning Analytics

Mina Shirvani Boroujeni
Nov. 08 2016

MOOCs Components

Videos

Forum

Problems

Logged events

Videos

Event Types

Video.Load

Video.Play

Video.Download

Video.Seek

Video.Pause

Video.SpeedChange

Video.Error

Video.Stalled

Forum

Event Types

Forum.Load

Forum.Subscribe

Forum.ThreadSubscribe

Forum.Thread.Launch

Forum.Thread.View

Forum.Thread.PostOn

Forum.Post.Upvote

Forum.Post.Downvote

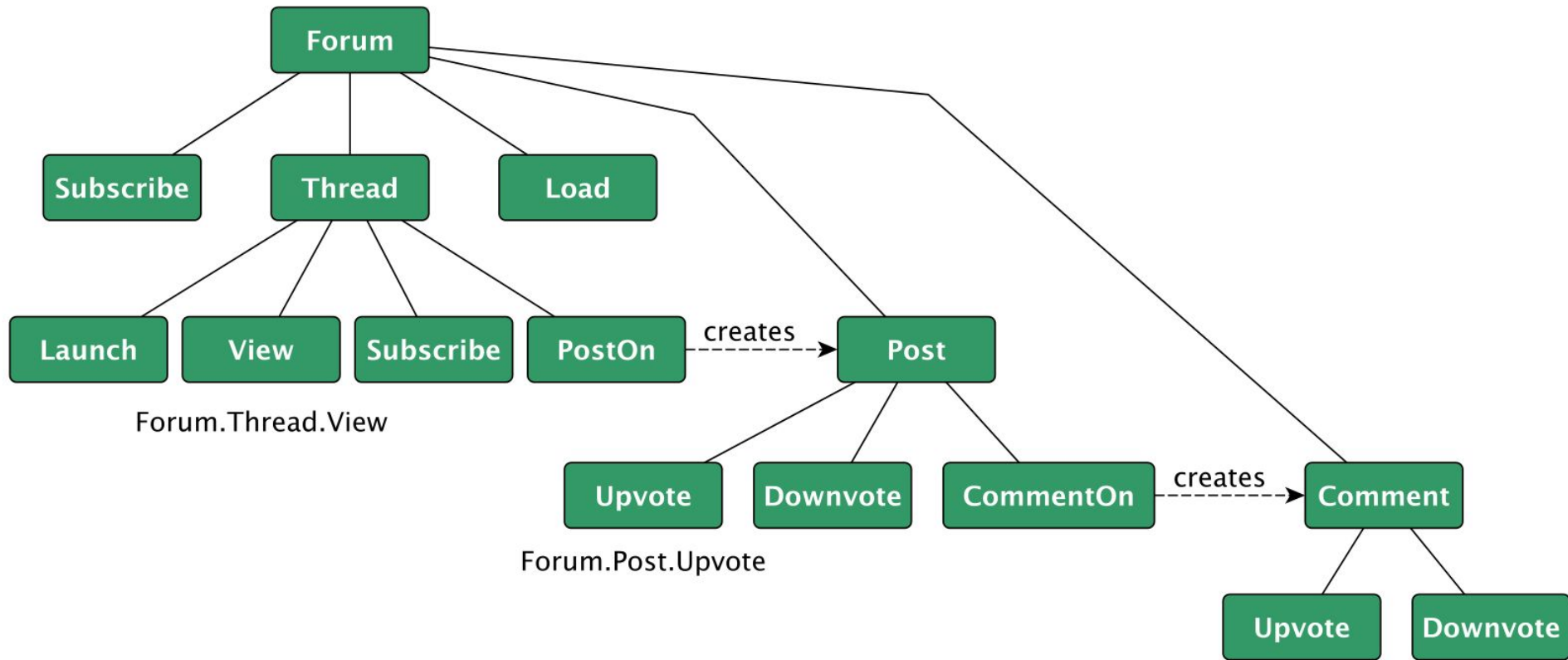
Forum.Comment.Downvote

Problems

Event Types

Problem.Check

Forum Hierarchy



Project Overview

- Assignment submission
 - Multiple submissions are possible
 - Different strategies followed by students:
 - Trial and error (bursty resubmission sequences)
 - Review course materials and resubmit
 - Review discussion forum and resubmit
 - Pose a question in discussion forum and resubmit
 - etc
- Overall goal:
 - Explore assignment submission strategies
 - Predict if the grade would improve after a resubmission to an assignment
 - Predict grade improvement between the first and last submission to an assignment

Data Analysis Pipeline

- Develop a hypothesis
- Extract features
- Partition data into train and test set
- Train model (on train set)
- Evaluate model performance (on test set)
- Find the best feature set and model

Model training and parameter selection in CARET

```
1 Define sets of model parameter values to evaluate
2 for each parameter set do
3   for each resampling iteration do
4     Hold-out specific samples
5     [Optional] Pre-process the data
6     Fit the model on the remainder
7     Predict the hold-out samples
8   end
9   Calculate the average performance across hold-out predictions
10 end
11 Determine the optimal parameter set
12 Fit the final model to all the training data using the optimal parameter set
```

Model training and parameter selection in CARET

- **Train**

```
ctrl= trainControl(method = 'cv')  
paramGrid <- expand.grid(C = c(0.001,0.01,0.1,0.5 ,1, 2, 3, 4))
```

```
model=train(x=d.train[,fs],  
            y=d.train$Price,  
            method = "svmLinear",  
            trControl = ctrl,  
            tuneGrid = paramGrid,  
            preProc = c("center", "scale") )
```

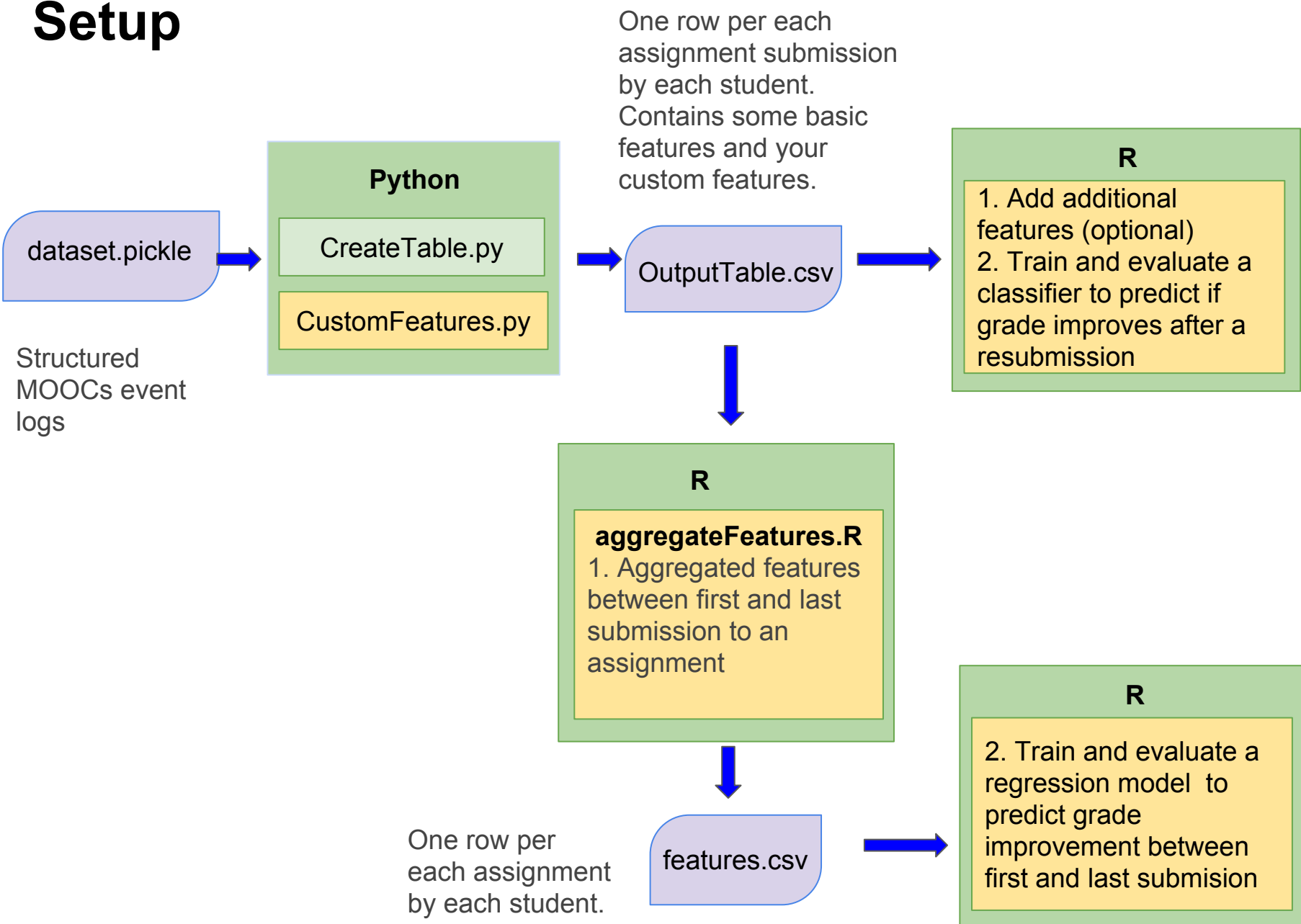
- **Test**

```
predict(model, newdata=d.test)
```


A few notes

- Develop features and build predictive model using the dataset provided on Moodle
- Find the best feature set and the best model
- For the final submission, you will use your model to
- Avoid including features based on assignment grades (grades would not be provided in the final test data)

Setup

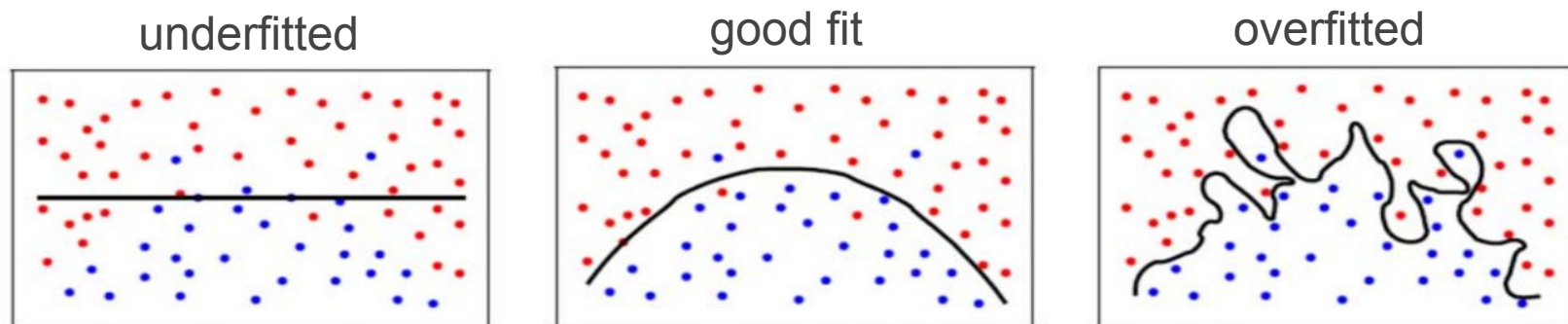


To get started with the R script

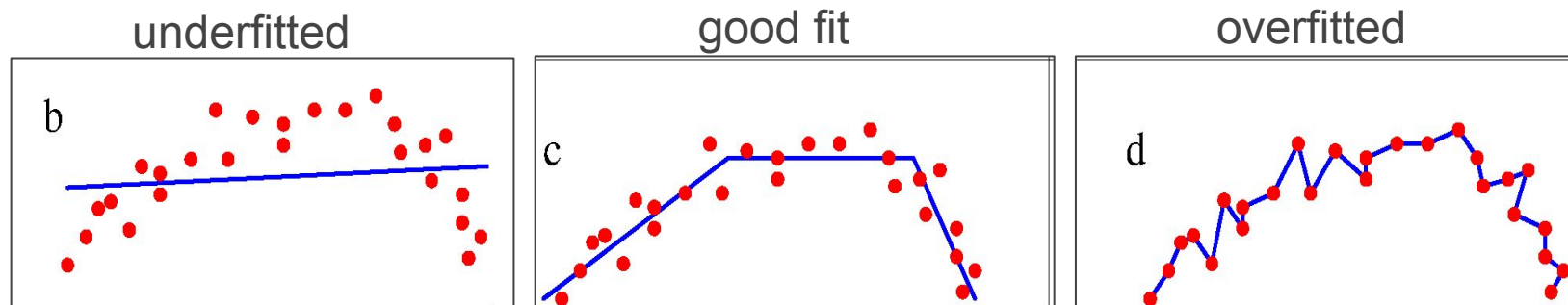
- Download the data and scripts from Moodle
- Modify '*aggregateFeatures.R*' and add a new feature encoding average time between resubmissions *avgTbwSubs*
- Train and SVM model to predict *overalGradeDiff* given *avgTbwSubs* and *countOfVideoandForumEvents*

Trade off: goodness of fit and generalizability

- Classification



- Regression



Cross Validation

A model validation technique for assessing how well the model generalizes to new data.

- Split training data into **training** and **validation** set
- Build the models using training data
- Evaluate the model on validation data