

XGB per-Tissue training and model evaluation

Required package

```
In [39]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from IPython.display import display, HTML
from pathlib import Path
from os import listdir
import sklearn.model_selection
import sklearn.feature_selection
%matplotlib inline
sns.set_style("whitegrid")

import xgboost as xgb
from xgboost.sklearn import XGBClassifier
from sklearn.model_selection import cross_validate, GridSearchCV
from sklearn import metrics #Additional sklearn functions
```

Supporting functions

```
In [7]: def predOneTissue(alg, dtrain, y_train, useTrainCV=True, cvFolds=5, early_stopping_rounds=50):
    if useTrainCV:
        xgb_param = alg.get_xgb_params()
        extra = {'num_class': len(y_train.unique())}
        xgbParams = alg.get_xgb_params()
        xgbParams.update(extra)
        xgTrain = xgb.DMatrix(dtrain, label=y_train)
        cvresult = xgb.cv(xgbParams,
                           xgTrain,
                           num_boost_round=alg.get_params()['n_estimators']
                           ,
                           nfold=cvFolds,
                           stratified=True,
                           metrics={'mlogloss'},
                           early_stopping_rounds=early_stopping_rounds,
                           seed=0,
                           callbacks=[xgb.callback.print_evaluation(show_st
```

```

dv=False),

                                xgb.callback.early_stop(3)])

    #print(cvresult)
    alg.set_params(n_estimators=cvresult.shape[0])
    alg.set_params(num_class=len(y_train.unique()))
    #Fit alg
    print(dtrain.shape,y_train.shape)
    alg.fit(dtrain,y_train,eval_metric='auc')
    return alg
def scoreOneTissue(y_test,y_preds):
    """
    Returns a tuple of raw score and F1.
    """

    return (metrics.accuracy_score(y_test,y_preds))
def splitData(X,y,rng=1234):
    return sklearn.model_selection.train_test_split(X, y, test_size=.3
, random_state=rng)
def simpleExpressionFilter(counts,min_count):
    """accepts raw counts and a minimum sum count per gene across all
samples
    return a boolean array of all genes, which can be applied to any t
ransformed counts.
    True is associated with passing the test.
    """

    keep=np.sum(counts)>min_count
    #print("Pre",counts.shape[1])
    #filtered_counts=counts.loc[:,(keep)] # similar to how the boolean
array would be used on any count matrix
    #print("Post",filtered_counts.shape[1])
    return(keep)
def loadData(TISSUE,data_dir,tissue_dir):
    cpm=pd.read_csv(data_dir/tissue_dir/str(TISSUE+"_cpm.tsv"),sep="\t
",index_col=0)
    #lcpm=pd.read_csv(data_dir/tissue_dir/str(TISSUE+"_lcpm.tsv"),sep=
"\t",index_col=0)
    cdat=pd.read_csv(data_dir/tissue_dir/str(TISSUE+"_c.tsv"),sep="\t"
,index_col=0)
    return (cpm,cdat)

```

Preparing metadata

```
In [9]: data_dir=Path("data")
tissue_dir=Path("tissue-specific")
manifest={"data":"All_Tissue_Site_Details.combined.reads.gct",
          "sample_meta":"GTEx_v7_Annotations_SampleAttributesDS.txt",
          "subject_meta":"GTEx_v7_Annotations_SubjectPhenotypesDS.txt",
          "merged_meta":"merged_meta.tsv"}
meta=pd.read_csv(data_dir/manifest['merged_meta'],sep="\t",dtype={'SMU
BRID':object,'SEX':object,'DTHHRDY':object})
meta=meta[~(meta['AGE'].isnull())]
counts=pd.DataFrame(meta['SMTS'].value_counts())
df=meta[meta['SMTS'].isin(counts[counts['SMTS']>200].index)]
```

Main Loop

```
In [ ]: def mainLoop(df,data_dir,tissue_dir):
    data_dir=Path("data")
    tissue_dir=Path("tissue-specific")
    results=[]
    algs=[]
    TISSUE_list=df['SMTS'].unique()
    for TISSUE in TISSUE_list:
        cpm,cdat=loadData(TISSUE,data_dir,tissue_dir)
        #print(cpm.shape)
        #print(cdat.shape)
        all_age=df.loc[(df['SMTS']==TISSUE),'AGE']
        #print(all_age.shape)
        cpm_train,cpm_test,y_train,y_test=splitData(cpm,all_age)
        c_train,c_test,y_train,y_test=splitData(cdat,all_age)
        #y_train.map({'20-29':0,'30-39':1,'40-49':2,'50-59':3,'60-69':4,'70-79':5})
        #y_test.map({'20-29':0,'30-39':1,'40-49':2,'50-59':3,'60-69':4,'70-79':5})
        y_train.replace({"20-29":0, "30-39":1, "40-49":2, "50-59":3, "60-69":4, "70-79":5}, inplace = True)
        y_test.replace({"20-29":0, "30-39":1, "40-49":2, "50-59":3, "60-69":4, "70-79":5}, inplace = True)
        #print(y_train.shape)
        #print(cpm_train.shape)
        #print(c_train.shape)
        keep=simpleExpressionFilter(c_train,10)
        cpm_train=cpm_train.loc[:,(keep)]
        cpm_test=cpm_test.loc[:,(keep)]

        selector=sklearn.feature_selection.VarianceThreshold(threshold=.1)
```

```

selector.fit(cpm_train)
keep=selector.get_support(indices=True)
cpm_train=cpm_train.iloc[:,keep]
cpm_test=cpm_test.iloc[:,keep]

xgb1 = XGBClassifier(
    learning_rate =0.1,
    n_estimators=1000,
    max_depth=5,
    min_child_weight=1,
    gamma=0,
    subsample=0.8,
    colsample_bytree=0.8,
    objective= 'multi:softprob',
    nthread=30,
    scale_pos_weight=1,
    seed=1234)
xgb1=predOneTissue(xgb1,cpm_train,y_train)
y_preds=xgb1.predict(cpm_test)
score=scoreOneTissue(y_test,y_preds)
#print(y_test,y_preds)
results.append(score)
algs.append(xgb1)
results=pd.Series(results,index=TISSUE_list)
return (algs,results)

```

Generating Models

In [12]: `algs,results=mainLoop(df,data_dir,tissue_dir)`

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precatd and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precatd and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```
[0]      train-mlogloss:1.62723  test-mlogloss:1.75625
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.49117  test-mlogloss:1.71253
[2]      train-mlogloss:1.37144  test-mlogloss:1.68676
[3]      train-mlogloss:1.26631  test-mlogloss:1.65759
[4]      train-mlogloss:1.17219  test-mlogloss:1.63195
[5]      train-mlogloss:1.08398  test-mlogloss:1.61364
[6]      train-mlogloss:1.00558  test-mlogloss:1.59863
[7]      train-mlogloss:0.936582 test-mlogloss:1.57958
[8]      train-mlogloss:0.872804 test-mlogloss:1.56196
[9]      train-mlogloss:0.816358 test-mlogloss:1.54551
[10]     train-mlogloss:0.763223 test-mlogloss:1.53085
[11]     train-mlogloss:0.715494 test-mlogloss:1.51898
[12]     train-mlogloss:0.670301 test-mlogloss:1.50313
[13]     train-mlogloss:0.62889  test-mlogloss:1.4951
[14]     train-mlogloss:0.589712 test-mlogloss:1.48627
[15]     train-mlogloss:0.553587 test-mlogloss:1.48002
[16]     train-mlogloss:0.521056 test-mlogloss:1.47262
[17]     train-mlogloss:0.490608 test-mlogloss:1.46529
[18]     train-mlogloss:0.462679 test-mlogloss:1.459
[19]     train-mlogloss:0.43558  test-mlogloss:1.45243
[20]     train-mlogloss:0.410768 test-mlogloss:1.44806
[21]     train-mlogloss:0.387903 test-mlogloss:1.44222
[22]     train-mlogloss:0.366456 test-mlogloss:1.43905
[23]     train-mlogloss:0.345849 test-mlogloss:1.43566
[24]     train-mlogloss:0.326704 test-mlogloss:1.4329
[25]     train-mlogloss:0.309109 test-mlogloss:1.42931
[26]     train-mlogloss:0.292335 test-mlogloss:1.42535
[27]     train-mlogloss:0.277248 test-mlogloss:1.42256
[28]     train-mlogloss:0.262898 test-mlogloss:1.42048
[29]     train-mlogloss:0.249399 test-mlogloss:1.42165
[30]     train-mlogloss:0.236063 test-mlogloss:1.42017
[31]     train-mlogloss:0.224707 test-mlogloss:1.41965
[32]     train-mlogloss:0.213066 test-mlogloss:1.41771
[33]     train-mlogloss:0.202521 test-mlogloss:1.41786
[34]     train-mlogloss:0.19248  test-mlogloss:1.4189
[35]     train-mlogloss:0.182935 test-mlogloss:1.41509
[36]     train-mlogloss:0.174465 test-mlogloss:1.41554
[37]     train-mlogloss:0.165938 test-mlogloss:1.41677
[38]     train-mlogloss:0.158326 test-mlogloss:1.41678
```

Stopping. Best iteration:

```
[35]      train-mlogloss:0.182935+0.00158929      test-mlogloss:1.4150
9+0.0303085
```

```
(557, 18506) (557,)
```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```

[0]      train-mlogloss:1.6308    test-mlogloss:1.74831
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1]      train-mlogloss:1.49492    test-mlogloss:1.70418
[2]      train-mlogloss:1.37817    test-mlogloss:1.66758
[3]      train-mlogloss:1.27171    test-mlogloss:1.63753
[4]      train-mlogloss:1.17936    test-mlogloss:1.61212
[5]      train-mlogloss:1.09242    test-mlogloss:1.58946
[6]      train-mlogloss:1.01204    test-mlogloss:1.56559
[7]      train-mlogloss:0.943112   test-mlogloss:1.54294
[8]      train-mlogloss:0.879043   test-mlogloss:1.52504
[9]      train-mlogloss:0.820085   test-mlogloss:1.50941
[10]     train-mlogloss:0.766329   test-mlogloss:1.49194
[11]     train-mlogloss:0.716385   test-mlogloss:1.4783
[12]     train-mlogloss:0.6714     test-mlogloss:1.46358
[13]     train-mlogloss:0.629      test-mlogloss:1.45327
[14]     train-mlogloss:0.590004   test-mlogloss:1.44447
[15]     train-mlogloss:0.555076   test-mlogloss:1.43161
[16]     train-mlogloss:0.522489   test-mlogloss:1.42111
[17]     train-mlogloss:0.491921   test-mlogloss:1.41204
[18]     train-mlogloss:0.463489   test-mlogloss:1.40248
[19]     train-mlogloss:0.437538   test-mlogloss:1.39416
[20]     train-mlogloss:0.412804   test-mlogloss:1.38831
[21]     train-mlogloss:0.390834   test-mlogloss:1.38183
[22]     train-mlogloss:0.368833   test-mlogloss:1.37436
[23]     train-mlogloss:0.348625   test-mlogloss:1.36611
[24]     train-mlogloss:0.330424   test-mlogloss:1.36071
[25]     train-mlogloss:0.312597   test-mlogloss:1.35603
[26]     train-mlogloss:0.296414   test-mlogloss:1.35028
[27]     train-mlogloss:0.280593   test-mlogloss:1.3484
[28]     train-mlogloss:0.266667   test-mlogloss:1.34369
[29]     train-mlogloss:0.253221   test-mlogloss:1.33806
[30]     train-mlogloss:0.240454   test-mlogloss:1.33358
[31]     train-mlogloss:0.228715   test-mlogloss:1.32969
[32]     train-mlogloss:0.217666   test-mlogloss:1.32541
[33]     train-mlogloss:0.20695    test-mlogloss:1.32332
[34]     train-mlogloss:0.197191   test-mlogloss:1.32198
[35]     train-mlogloss:0.187614   test-mlogloss:1.31899
[36]     train-mlogloss:0.179081   test-mlogloss:1.31531

```

```

[37] train-mlogloss:0.170674 test-mlogloss:1.31414
[38] train-mlogloss:0.162475 test-mlogloss:1.31313
[39] train-mlogloss:0.154937 test-mlogloss:1.30864
[40] train-mlogloss:0.147806 test-mlogloss:1.30566
[41] train-mlogloss:0.140819 test-mlogloss:1.30128
[42] train-mlogloss:0.134382 test-mlogloss:1.30006
[43] train-mlogloss:0.128308 test-mlogloss:1.29918
[44] train-mlogloss:0.122616 test-mlogloss:1.29908
[45] train-mlogloss:0.117377 test-mlogloss:1.29899
[46] train-mlogloss:0.112361 test-mlogloss:1.29683
[47] train-mlogloss:0.107609 test-mlogloss:1.29664
[48] train-mlogloss:0.103365 test-mlogloss:1.298
[49] train-mlogloss:0.0989942 test-mlogloss:1.29713
[50] train-mlogloss:0.0951044 test-mlogloss:1.29567
[51] train-mlogloss:0.0914834 test-mlogloss:1.29596
[52] train-mlogloss:0.0879578 test-mlogloss:1.29379
[53] train-mlogloss:0.0845932 test-mlogloss:1.29345
[54] train-mlogloss:0.0814428 test-mlogloss:1.29215
[55] train-mlogloss:0.0785142 test-mlogloss:1.29153
[56] train-mlogloss:0.0756948 test-mlogloss:1.29262
[57] train-mlogloss:0.0731404 test-mlogloss:1.29109
[58] train-mlogloss:0.070635 test-mlogloss:1.29215
[59] train-mlogloss:0.068195 test-mlogloss:1.29172
[60] train-mlogloss:0.0658674 test-mlogloss:1.29205
Stopping. Best iteration:
[57] train-mlogloss:0.0731404+0.00113844 test-mlogloss:1.2910
9+0.065828

```

```
(639, 17997) (639,)
```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
  if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
  data.base is not None and isinstance(data, np.ndarray) \

```

```
[0]      train-mlogloss:1.60869  test-mlogloss:1.73007
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.45709  test-mlogloss:1.68186
[2]      train-mlogloss:1.32824  test-mlogloss:1.64892
[3]      train-mlogloss:1.21237  test-mlogloss:1.60797
[4]      train-mlogloss:1.11179  test-mlogloss:1.578
[5]      train-mlogloss:1.02635  test-mlogloss:1.55
[6]      train-mlogloss:0.945054 test-mlogloss:1.5256
[7]      train-mlogloss:0.873949 test-mlogloss:1.50352
[8]      train-mlogloss:0.809525 test-mlogloss:1.48985
[9]      train-mlogloss:0.751001 test-mlogloss:1.46971
[10]     train-mlogloss:0.697382 test-mlogloss:1.45134
[11]     train-mlogloss:0.649128 test-mlogloss:1.43551
[12]     train-mlogloss:0.603495 test-mlogloss:1.4226
[13]     train-mlogloss:0.56239  test-mlogloss:1.41392
[14]     train-mlogloss:0.52452  test-mlogloss:1.40615
[15]     train-mlogloss:0.489726 test-mlogloss:1.39553
[16]     train-mlogloss:0.45799  test-mlogloss:1.38812
[17]     train-mlogloss:0.428098 test-mlogloss:1.38105
[18]     train-mlogloss:0.400087 test-mlogloss:1.37241
[19]     train-mlogloss:0.374912 test-mlogloss:1.36779
[20]     train-mlogloss:0.351439 test-mlogloss:1.36071
[21]     train-mlogloss:0.328771 test-mlogloss:1.35794
[22]     train-mlogloss:0.30866  test-mlogloss:1.3527
[23]     train-mlogloss:0.29052  test-mlogloss:1.34996
[24]     train-mlogloss:0.273703 test-mlogloss:1.34662
[25]     train-mlogloss:0.257568 test-mlogloss:1.34478
[26]     train-mlogloss:0.243275 test-mlogloss:1.34093
[27]     train-mlogloss:0.229438 test-mlogloss:1.33605
[28]     train-mlogloss:0.216826 test-mlogloss:1.33442
[29]     train-mlogloss:0.204739 test-mlogloss:1.33096
[30]     train-mlogloss:0.193957 test-mlogloss:1.33368
[31]     train-mlogloss:0.183691 test-mlogloss:1.33213
[32]     train-mlogloss:0.174517 test-mlogloss:1.33426
```

Stopping. Best iteration:

```
[29]      train-mlogloss:0.204739+0.00405159      test-mlogloss:1.3309
6+0.0444539
```

```
(231, 18529) (231,)
```



```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```

[0]      train-mlogloss:1.62458  test-mlogloss:1.7603
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1]      train-mlogloss:1.48475  test-mlogloss:1.75029
[2]      train-mlogloss:1.35653  test-mlogloss:1.73464
[3]      train-mlogloss:1.24572  test-mlogloss:1.72469
[4]      train-mlogloss:1.1469   test-mlogloss:1.71252
[5]      train-mlogloss:1.05638  test-mlogloss:1.71053
[6]      train-mlogloss:0.977475 test-mlogloss:1.69693
[7]      train-mlogloss:0.904182 test-mlogloss:1.68281
[8]      train-mlogloss:0.839464 test-mlogloss:1.67911
[9]      train-mlogloss:0.780184 test-mlogloss:1.67069
[10]     train-mlogloss:0.725435 test-mlogloss:1.66413
[11]     train-mlogloss:0.676193 test-mlogloss:1.66201
[12]     train-mlogloss:0.629202 test-mlogloss:1.66187
[13]     train-mlogloss:0.588011 test-mlogloss:1.65833
[14]     train-mlogloss:0.549467 test-mlogloss:1.65554
[15]     train-mlogloss:0.513601 test-mlogloss:1.65407
[16]     train-mlogloss:0.480563 test-mlogloss:1.65677
[17]     train-mlogloss:0.450531 test-mlogloss:1.65371
[18]     train-mlogloss:0.422136 test-mlogloss:1.65354
[19]     train-mlogloss:0.39713  test-mlogloss:1.65294
[20]     train-mlogloss:0.373087 test-mlogloss:1.6523
[21]     train-mlogloss:0.351338 test-mlogloss:1.65349
[22]     train-mlogloss:0.330303 test-mlogloss:1.6562
[23]     train-mlogloss:0.311714 test-mlogloss:1.66023

```

Stopping. Best iteration:

```

[20]      train-mlogloss:0.373087+0.00276797      test-mlogloss:1.6523
+0.0484465

```

```

(203, 18551) (203,)

```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```

[0]      train-mlogloss:1.64027  test-mlogloss:1.76577
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1]      train-mlogloss:1.4993  test-mlogloss:1.74103
[2]      train-mlogloss:1.38147  test-mlogloss:1.72017
[3]      train-mlogloss:1.27506  test-mlogloss:1.70215
[4]      train-mlogloss:1.17606  test-mlogloss:1.6904
[5]      train-mlogloss:1.08882  test-mlogloss:1.67612
[6]      train-mlogloss:1.01018  test-mlogloss:1.6666
[7]      train-mlogloss:0.938717 test-mlogloss:1.65965
[8]      train-mlogloss:0.870944 test-mlogloss:1.64999
[9]      train-mlogloss:0.812613 test-mlogloss:1.64066
[10]     train-mlogloss:0.758909 test-mlogloss:1.63313
[11]     train-mlogloss:0.708976 test-mlogloss:1.62378
[12]     train-mlogloss:0.663855 test-mlogloss:1.61797
[13]     train-mlogloss:0.623118 test-mlogloss:1.6103
[14]     train-mlogloss:0.584143 test-mlogloss:1.60018
[15]     train-mlogloss:0.548921 test-mlogloss:1.59729
[16]     train-mlogloss:0.516251 test-mlogloss:1.5943
[17]     train-mlogloss:0.486023 test-mlogloss:1.59024
[18]     train-mlogloss:0.456896 test-mlogloss:1.5928
[19]     train-mlogloss:0.429269 test-mlogloss:1.59147
[20]     train-mlogloss:0.403862 test-mlogloss:1.59057

```

Stopping. Best iteration:

```

[17]      train-mlogloss:0.486023+0.00426828      test-mlogloss:1.5902
4+0.0493802

```

```

(375, 19146) (375,)

```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```

[0]      train-mlogloss:1.65942  test-mlogloss:1.7606

```

Multiple eval metrics have been passed: 'test-mlogloss' will be used for early stopping.

Will train until test-mlogloss hasn't improved in 3 rounds.

[1]	train-mlogloss:1.5461	test-mlogloss:1.73419
[2]	train-mlogloss:1.44427	test-mlogloss:1.71035
[3]	train-mlogloss:1.34472	test-mlogloss:1.68839
[4]	train-mlogloss:1.25756	test-mlogloss:1.67214
[5]	train-mlogloss:1.18018	test-mlogloss:1.65544
[6]	train-mlogloss:1.10842	test-mlogloss:1.6387
[7]	train-mlogloss:1.0411	test-mlogloss:1.62583
[8]	train-mlogloss:0.979213	test-mlogloss:1.61107
[9]	train-mlogloss:0.922133	test-mlogloss:1.60064
[10]	train-mlogloss:0.867716	test-mlogloss:1.59166
[11]	train-mlogloss:0.818085	test-mlogloss:1.58238
[12]	train-mlogloss:0.772857	test-mlogloss:1.57112
[13]	train-mlogloss:0.729558	test-mlogloss:1.56237
[14]	train-mlogloss:0.690405	test-mlogloss:1.55529
[15]	train-mlogloss:0.654781	test-mlogloss:1.5493
[16]	train-mlogloss:0.619542	test-mlogloss:1.54388
[17]	train-mlogloss:0.586648	test-mlogloss:1.53871
[18]	train-mlogloss:0.555719	test-mlogloss:1.5332
[19]	train-mlogloss:0.52638	test-mlogloss:1.52706
[20]	train-mlogloss:0.499203	test-mlogloss:1.52311
[21]	train-mlogloss:0.47395	test-mlogloss:1.51876
[22]	train-mlogloss:0.449415	test-mlogloss:1.51544
[23]	train-mlogloss:0.427799	test-mlogloss:1.5112
[24]	train-mlogloss:0.407205	test-mlogloss:1.5093
[25]	train-mlogloss:0.388291	test-mlogloss:1.50795
[26]	train-mlogloss:0.370158	test-mlogloss:1.5045
[27]	train-mlogloss:0.352208	test-mlogloss:1.50141
[28]	train-mlogloss:0.335585	test-mlogloss:1.50088
[29]	train-mlogloss:0.320115	test-mlogloss:1.49924
[30]	train-mlogloss:0.30528	test-mlogloss:1.49704
[31]	train-mlogloss:0.292292	test-mlogloss:1.49612
[32]	train-mlogloss:0.279638	test-mlogloss:1.49591
[33]	train-mlogloss:0.266705	test-mlogloss:1.49548
[34]	train-mlogloss:0.254815	test-mlogloss:1.49727
[35]	train-mlogloss:0.243812	test-mlogloss:1.49611
[36]	train-mlogloss:0.233321	test-mlogloss:1.49363
[37]	train-mlogloss:0.22297	test-mlogloss:1.49288
[38]	train-mlogloss:0.213087	test-mlogloss:1.493
[39]	train-mlogloss:0.204229	test-mlogloss:1.49477
[40]	train-mlogloss:0.195689	test-mlogloss:1.49467

Stopping. Best iteration:

[37]	train-mlogloss:0.22297+0.00469227	test-mlogloss:1.49288+0.0154897
------	-----------------------------------	---------------------------------

(841, 18824) (841,)

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
recated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
recated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.63103  test-mlogloss:1.75599
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.48559  test-mlogloss:1.71649
[2]      train-mlogloss:1.36021  test-mlogloss:1.69182
[3]      train-mlogloss:1.24793  test-mlogloss:1.66629
[4]      train-mlogloss:1.14974  test-mlogloss:1.64098
[5]      train-mlogloss:1.05818  test-mlogloss:1.62059
[6]      train-mlogloss:0.978814 test-mlogloss:1.60472
[7]      train-mlogloss:0.907214 test-mlogloss:1.58399
[8]      train-mlogloss:0.839631 test-mlogloss:1.5727
[9]      train-mlogloss:0.778291 test-mlogloss:1.56026
[10]     train-mlogloss:0.725874 test-mlogloss:1.54591
[11]     train-mlogloss:0.677309 test-mlogloss:1.53461
[12]     train-mlogloss:0.630819 test-mlogloss:1.52624
[13]     train-mlogloss:0.589918 test-mlogloss:1.5181
[14]     train-mlogloss:0.551226 test-mlogloss:1.50802
[15]     train-mlogloss:0.516764 test-mlogloss:1.50194
[16]     train-mlogloss:0.484652 test-mlogloss:1.50153
[17]     train-mlogloss:0.454063 test-mlogloss:1.49743
[18]     train-mlogloss:0.426658 test-mlogloss:1.49511
[19]     train-mlogloss:0.400681 test-mlogloss:1.49048
[20]     train-mlogloss:0.376167 test-mlogloss:1.48677
[21]     train-mlogloss:0.353979 test-mlogloss:1.48253
[22]     train-mlogloss:0.333599 test-mlogloss:1.47787
[23]     train-mlogloss:0.314118 test-mlogloss:1.47738
[24]     train-mlogloss:0.296512 test-mlogloss:1.47727
[25]     train-mlogloss:0.279499 test-mlogloss:1.47436
[26]     train-mlogloss:0.263937 test-mlogloss:1.47349
[27]     train-mlogloss:0.249454 test-mlogloss:1.47112
[28]     train-mlogloss:0.235844 test-mlogloss:1.47055
[29]     train-mlogloss:0.223135 test-mlogloss:1.46951
[30]     train-mlogloss:0.211588 test-mlogloss:1.46806
[31]     train-mlogloss:0.200658 test-mlogloss:1.4678
[32]     train-mlogloss:0.190662 test-mlogloss:1.46786
[33]     train-mlogloss:0.181013 test-mlogloss:1.46599
[34]     train-mlogloss:0.172453 test-mlogloss:1.46661
[35]     train-mlogloss:0.164287 test-mlogloss:1.46429
[36]     train-mlogloss:0.156439 test-mlogloss:1.46466
[37]     train-mlogloss:0.149155 test-mlogloss:1.46562
[38]     train-mlogloss:0.142424 test-mlogloss:1.46434
```

Stopping. Best iteration:

```
[35]      train-mlogloss:0.164287+0.00319133      test-mlogloss:1.4642
9+0.0485674
```

```
(354, 18572) (354,)
```

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
  if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
  data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.64576  test-mlogloss:1.75233
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.51872  test-mlogloss:1.71793
[2]      train-mlogloss:1.40329  test-mlogloss:1.68776
[3]      train-mlogloss:1.2978   test-mlogloss:1.66108
[4]      train-mlogloss:1.20616  test-mlogloss:1.64237
[5]      train-mlogloss:1.12177  test-mlogloss:1.62323
[6]      train-mlogloss:1.04434  test-mlogloss:1.60632
[7]      train-mlogloss:0.974897 test-mlogloss:1.58708
[8]      train-mlogloss:0.911448 test-mlogloss:1.57222
[9]      train-mlogloss:0.854488 test-mlogloss:1.5573
[10]     train-mlogloss:0.802543 test-mlogloss:1.54635
[11]     train-mlogloss:0.753999 test-mlogloss:1.53584
[12]     train-mlogloss:0.707438 test-mlogloss:1.52743
[13]     train-mlogloss:0.663252 test-mlogloss:1.51648
[14]     train-mlogloss:0.624073 test-mlogloss:1.5074
[15]     train-mlogloss:0.587505 test-mlogloss:1.49742
[16]     train-mlogloss:0.554091 test-mlogloss:1.4914
[17]     train-mlogloss:0.52317  test-mlogloss:1.48284
[18]     train-mlogloss:0.492856 test-mlogloss:1.47422
[19]     train-mlogloss:0.466324 test-mlogloss:1.46619
[20]     train-mlogloss:0.440813 test-mlogloss:1.45887
[21]     train-mlogloss:0.417486 test-mlogloss:1.45318
[22]     train-mlogloss:0.396002 test-mlogloss:1.44681
[23]     train-mlogloss:0.374873 test-mlogloss:1.44085
[24]     train-mlogloss:0.35496  test-mlogloss:1.43746
[25]     train-mlogloss:0.336812 test-mlogloss:1.43301
[26]     train-mlogloss:0.320136 test-mlogloss:1.42998
[27]     train-mlogloss:0.303891 test-mlogloss:1.42615
[28]     train-mlogloss:0.289074 test-mlogloss:1.42304
[29]     train-mlogloss:0.275324 test-mlogloss:1.42247
[30]     train-mlogloss:0.262113 test-mlogloss:1.41945
[31]     train-mlogloss:0.249272 test-mlogloss:1.41533
[32]     train-mlogloss:0.236873 test-mlogloss:1.41227
[33]     train-mlogloss:0.225781 test-mlogloss:1.41191
[34]     train-mlogloss:0.215278 test-mlogloss:1.41166
[35]     train-mlogloss:0.205218 test-mlogloss:1.40876
[36]     train-mlogloss:0.195773 test-mlogloss:1.4081
```

```

[37] train-mlogloss:0.187057 test-mlogloss:1.40626
[38] train-mlogloss:0.178751 test-mlogloss:1.40387
[39] train-mlogloss:0.17099 test-mlogloss:1.40335
[40] train-mlogloss:0.163524 test-mlogloss:1.40077
[41] train-mlogloss:0.15659 test-mlogloss:1.39925
[42] train-mlogloss:0.149882 test-mlogloss:1.39924
[43] train-mlogloss:0.143795 test-mlogloss:1.4001
[44] train-mlogloss:0.137877 test-mlogloss:1.39868
[45] train-mlogloss:0.132328 test-mlogloss:1.39579
[46] train-mlogloss:0.126993 test-mlogloss:1.39305
[47] train-mlogloss:0.121949 test-mlogloss:1.39202
[48] train-mlogloss:0.117135 test-mlogloss:1.39176
[49] train-mlogloss:0.112398 test-mlogloss:1.39107
[50] train-mlogloss:0.108086 test-mlogloss:1.39002
[51] train-mlogloss:0.103869 test-mlogloss:1.39212
[52] train-mlogloss:0.0999766 test-mlogloss:1.39451
[53] train-mlogloss:0.0963804 test-mlogloss:1.39307
Stopping. Best iteration:
[50] train-mlogloss:0.108086+0.00124693 test-mlogloss:1.3900
2+0.0740522

```

```
(714, 18679) (714,)
```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
  if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
  data.base is not None and isinstance(data, np.ndarray) \

```

```

[0] train-mlogloss:1.62562 test-mlogloss:1.73792
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1] train-mlogloss:1.48387 test-mlogloss:1.69874
[2] train-mlogloss:1.36229 test-mlogloss:1.66588
[3] train-mlogloss:1.25312 test-mlogloss:1.63387
[4] train-mlogloss:1.15503 test-mlogloss:1.60495
[5] train-mlogloss:1.06851 test-mlogloss:1.5855
[6] train-mlogloss:0.989282 test-mlogloss:1.56055
[7] train-mlogloss:0.916157 test-mlogloss:1.54497
[8] train-mlogloss:0.851742 test-mlogloss:1.53047
[9] train-mlogloss:0.791991 test-mlogloss:1.5184
[10] train-mlogloss:0.739237 test-mlogloss:1.50638
[11] train-mlogloss:0.690256 test-mlogloss:1.493
[12] train-mlogloss:0.645892 test-mlogloss:1.48141
[13] train-mlogloss:0.603253 test-mlogloss:1.47381

```

```
[14] train-mlogloss:0.566304 test-mlogloss:1.46186
[15] train-mlogloss:0.530764 test-mlogloss:1.45148
[16] train-mlogloss:0.497048 test-mlogloss:1.44029
[17] train-mlogloss:0.465624 test-mlogloss:1.43043
[18] train-mlogloss:0.437096 test-mlogloss:1.42462
[19] train-mlogloss:0.411353 test-mlogloss:1.41914
[20] train-mlogloss:0.385403 test-mlogloss:1.41758
[21] train-mlogloss:0.362504 test-mlogloss:1.41063
[22] train-mlogloss:0.341547 test-mlogloss:1.40675
[23] train-mlogloss:0.321808 test-mlogloss:1.40019
[24] train-mlogloss:0.303203 test-mlogloss:1.39738
[25] train-mlogloss:0.286359 test-mlogloss:1.39478
[26] train-mlogloss:0.270438 test-mlogloss:1.39237
[27] train-mlogloss:0.255304 test-mlogloss:1.39023
[28] train-mlogloss:0.241404 test-mlogloss:1.38751
[29] train-mlogloss:0.228163 test-mlogloss:1.38672
[30] train-mlogloss:0.216398 test-mlogloss:1.38564
[31] train-mlogloss:0.205123 test-mlogloss:1.38404
[32] train-mlogloss:0.194406 test-mlogloss:1.38115
[33] train-mlogloss:0.184399 test-mlogloss:1.38294
[34] train-mlogloss:0.174994 test-mlogloss:1.38024
[35] train-mlogloss:0.166296 test-mlogloss:1.37847
[36] train-mlogloss:0.158111 test-mlogloss:1.37747
[37] train-mlogloss:0.150521 test-mlogloss:1.37607
[38] train-mlogloss:0.143387 test-mlogloss:1.37538
[39] train-mlogloss:0.136689 test-mlogloss:1.3744
[40] train-mlogloss:0.130552 test-mlogloss:1.37425
[41] train-mlogloss:0.124793 test-mlogloss:1.37274
[42] train-mlogloss:0.119169 test-mlogloss:1.37331
[43] train-mlogloss:0.11385 test-mlogloss:1.37133
[44] train-mlogloss:0.109112 test-mlogloss:1.36898
[45] train-mlogloss:0.104461 test-mlogloss:1.36795
[46] train-mlogloss:0.100214 test-mlogloss:1.36871
[47] train-mlogloss:0.0961874 test-mlogloss:1.3676
[48] train-mlogloss:0.0922636 test-mlogloss:1.36879
[49] train-mlogloss:0.0886658 test-mlogloss:1.37017
[50] train-mlogloss:0.0852774 test-mlogloss:1.37006
Stopping. Best iteration:
[47] train-mlogloss:0.0961874+0.00147797 test-mlogloss:1.3676
+0.0514605

(420, 18205) (420,)
```



```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.62618  test-mlogloss:1.75689
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

```
Will train until test-mlogloss hasn't improved in 3 rounds.
```

```
[1]      train-mlogloss:1.48062  test-mlogloss:1.72226
[2]      train-mlogloss:1.35458  test-mlogloss:1.6985
[3]      train-mlogloss:1.244    test-mlogloss:1.66294
[4]      train-mlogloss:1.14643  test-mlogloss:1.64615
[5]      train-mlogloss:1.05598  test-mlogloss:1.62481
[6]      train-mlogloss:0.976105 test-mlogloss:1.60881
[7]      train-mlogloss:0.904506 test-mlogloss:1.59394
[8]      train-mlogloss:0.838625 test-mlogloss:1.58139
[9]      train-mlogloss:0.778943 test-mlogloss:1.56729
[10]     train-mlogloss:0.724447 test-mlogloss:1.55246
[11]     train-mlogloss:0.675468 test-mlogloss:1.54068
[12]     train-mlogloss:0.63141  test-mlogloss:1.52983
[13]     train-mlogloss:0.590379 test-mlogloss:1.52442
[14]     train-mlogloss:0.552812 test-mlogloss:1.51944
[15]     train-mlogloss:0.517025 test-mlogloss:1.51271
[16]     train-mlogloss:0.485194 test-mlogloss:1.50755
[17]     train-mlogloss:0.455279 test-mlogloss:1.50627
[18]     train-mlogloss:0.428309 test-mlogloss:1.49983
[19]     train-mlogloss:0.403074 test-mlogloss:1.49848
[20]     train-mlogloss:0.377695 test-mlogloss:1.49346
[21]     train-mlogloss:0.354584 test-mlogloss:1.48909
[22]     train-mlogloss:0.33339  test-mlogloss:1.48123
[23]     train-mlogloss:0.314857 test-mlogloss:1.47844
[24]     train-mlogloss:0.297311 test-mlogloss:1.47374
[25]     train-mlogloss:0.280027 test-mlogloss:1.47344
[26]     train-mlogloss:0.263942 test-mlogloss:1.47384
[27]     train-mlogloss:0.249783 test-mlogloss:1.47322
[28]     train-mlogloss:0.23602  test-mlogloss:1.47391
[29]     train-mlogloss:0.223608 test-mlogloss:1.47133
[30]     train-mlogloss:0.21168  test-mlogloss:1.47084
[31]     train-mlogloss:0.200973 test-mlogloss:1.47074
[32]     train-mlogloss:0.190722 test-mlogloss:1.46967
[33]     train-mlogloss:0.181057 test-mlogloss:1.46796
[34]     train-mlogloss:0.172161 test-mlogloss:1.46808
[35]     train-mlogloss:0.163755 test-mlogloss:1.46999
[36]     train-mlogloss:0.156132 test-mlogloss:1.46958
```

```
Stopping. Best iteration:
```

```
[33]      train-mlogloss:0.181057+0.00336725      test-mlogloss:1.4679
6+0.0699253
```

```
(298, 18598) (298,)
```

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.63412  test-mlogloss:1.75296
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

```
Will train until test-mlogloss hasn't improved in 3 rounds.
```

```
[1]      train-mlogloss:1.48764  test-mlogloss:1.72757
[2]      train-mlogloss:1.36419  test-mlogloss:1.70735
[3]      train-mlogloss:1.24985  test-mlogloss:1.68651
[4]      train-mlogloss:1.1527   test-mlogloss:1.66528
[5]      train-mlogloss:1.06616  test-mlogloss:1.64322
[6]      train-mlogloss:0.98735  test-mlogloss:1.6318
[7]      train-mlogloss:0.914975 test-mlogloss:1.62502
[8]      train-mlogloss:0.85056  test-mlogloss:1.61506
[9]      train-mlogloss:0.79159  test-mlogloss:1.6042
[10]     train-mlogloss:0.738907 test-mlogloss:1.59265
[11]     train-mlogloss:0.689431 test-mlogloss:1.58432
[12]     train-mlogloss:0.645034 test-mlogloss:1.57394
[13]     train-mlogloss:0.602468 test-mlogloss:1.56765
[14]     train-mlogloss:0.564137 test-mlogloss:1.56108
[15]     train-mlogloss:0.527829 test-mlogloss:1.55526
[16]     train-mlogloss:0.494761 test-mlogloss:1.55028
[17]     train-mlogloss:0.464615 test-mlogloss:1.54311
[18]     train-mlogloss:0.437416 test-mlogloss:1.5347
[19]     train-mlogloss:0.411034 test-mlogloss:1.53229
[20]     train-mlogloss:0.386421 test-mlogloss:1.52835
[21]     train-mlogloss:0.363617 test-mlogloss:1.52277
[22]     train-mlogloss:0.34294   test-mlogloss:1.51689
[23]     train-mlogloss:0.323201 test-mlogloss:1.51341
[24]     train-mlogloss:0.304823 test-mlogloss:1.51162
[25]     train-mlogloss:0.287536 test-mlogloss:1.50918
[26]     train-mlogloss:0.271237 test-mlogloss:1.50835
[27]     train-mlogloss:0.25683   test-mlogloss:1.50666
[28]     train-mlogloss:0.243164 test-mlogloss:1.5089
[29]     train-mlogloss:0.230297 test-mlogloss:1.50653
[30]     train-mlogloss:0.217984 test-mlogloss:1.50503
[31]     train-mlogloss:0.207022 test-mlogloss:1.5026
[32]     train-mlogloss:0.196167 test-mlogloss:1.50092
[33]     train-mlogloss:0.186636 test-mlogloss:1.50017
[34]     train-mlogloss:0.177348 test-mlogloss:1.5011
[35]     train-mlogloss:0.168652 test-mlogloss:1.50231
[36]     train-mlogloss:0.160811 test-mlogloss:1.50383
```

```
Stopping. Best iteration:
```

```
[33]      train-mlogloss:0.186636+0.00407921      test-mlogloss:1.5001
7+0.0791445
```

```
(394, 17297) (394,)
```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \

```

```

[0]      train-mlogloss:1.62888  test-mlogloss:1.7531
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1]      train-mlogloss:1.48053  test-mlogloss:1.71451
[2]      train-mlogloss:1.35068  test-mlogloss:1.68514
[3]      train-mlogloss:1.23839  test-mlogloss:1.67023
[4]      train-mlogloss:1.13999  test-mlogloss:1.6503
[5]      train-mlogloss:1.04986  test-mlogloss:1.62632
[6]      train-mlogloss:0.968941 test-mlogloss:1.60814
[7]      train-mlogloss:0.898351 test-mlogloss:1.59219
[8]      train-mlogloss:0.831274 test-mlogloss:1.57617
[9]      train-mlogloss:0.774062 test-mlogloss:1.56629
[10]     train-mlogloss:0.721827 test-mlogloss:1.55429
[11]     train-mlogloss:0.67243  test-mlogloss:1.5396
[12]     train-mlogloss:0.627102 test-mlogloss:1.52808
[13]     train-mlogloss:0.585868 test-mlogloss:1.51681
[14]     train-mlogloss:0.548043 test-mlogloss:1.50622
[15]     train-mlogloss:0.513316 test-mlogloss:1.49849
[16]     train-mlogloss:0.480779 test-mlogloss:1.48679
[17]     train-mlogloss:0.450067 test-mlogloss:1.47802
[18]     train-mlogloss:0.422554 test-mlogloss:1.47099
[19]     train-mlogloss:0.396699 test-mlogloss:1.46435
[20]     train-mlogloss:0.373004 test-mlogloss:1.45413
[21]     train-mlogloss:0.35059  test-mlogloss:1.44913
[22]     train-mlogloss:0.329645 test-mlogloss:1.44477
[23]     train-mlogloss:0.310013 test-mlogloss:1.44011
[24]     train-mlogloss:0.292413 test-mlogloss:1.43379
[25]     train-mlogloss:0.275789 test-mlogloss:1.43033
[26]     train-mlogloss:0.260211 test-mlogloss:1.42761
[27]     train-mlogloss:0.246406 test-mlogloss:1.42271
[28]     train-mlogloss:0.232607 test-mlogloss:1.41976
[29]     train-mlogloss:0.220126 test-mlogloss:1.41715
[30]     train-mlogloss:0.208732 test-mlogloss:1.41383
[31]     train-mlogloss:0.197617 test-mlogloss:1.41072
[32]     train-mlogloss:0.18742  test-mlogloss:1.40937
[33]     train-mlogloss:0.177886 test-mlogloss:1.40734
[34]     train-mlogloss:0.169089 test-mlogloss:1.40779
[35]     train-mlogloss:0.160596 test-mlogloss:1.40736
[36]     train-mlogloss:0.152942 test-mlogloss:1.40623

```

```

[37]    train-mlogloss:0.145598 test-mlogloss:1.40682
[38]    train-mlogloss:0.138742 test-mlogloss:1.404
[39]    train-mlogloss:0.132636 test-mlogloss:1.40119
[40]    train-mlogloss:0.126778 test-mlogloss:1.40073
[41]    train-mlogloss:0.121179 test-mlogloss:1.39966
[42]    train-mlogloss:0.116059 test-mlogloss:1.4006
[43]    train-mlogloss:0.111205 test-mlogloss:1.3993
[44]    train-mlogloss:0.106597 test-mlogloss:1.39885
[45]    train-mlogloss:0.102319 test-mlogloss:1.39986
[46]    train-mlogloss:0.0983586      test-mlogloss:1.39966
[47]    train-mlogloss:0.0945034      test-mlogloss:1.4005
Stopping. Best iteration:
[44]    train-mlogloss:0.106597+0.0028279      test-mlogloss:1.3988
5+0.0549535

```

```
(289, 17483) (289,)
```

```

/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/sklearn/model_selection/_split.py:652: Warning: The 1
east populated class in y has only 1 members, which is too few. The
minimum number of members in any class cannot be less than n_splits=
5.
    % (min_groups, self.n_splits)), Warning)

```

```

[0]    train-mlogloss:1.626      test-mlogloss:1.74589
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.

```

Will train until test-mlogloss hasn't improved in 3 rounds.

```

[1]    train-mlogloss:1.48256 test-mlogloss:1.71312
[2]    train-mlogloss:1.35581 test-mlogloss:1.67524
[3]    train-mlogloss:1.24284 test-mlogloss:1.65572
[4]    train-mlogloss:1.14492 test-mlogloss:1.63376
[5]    train-mlogloss:1.05554 test-mlogloss:1.60612
[6]    train-mlogloss:0.973133 test-mlogloss:1.58753
[7]    train-mlogloss:0.897811 test-mlogloss:1.57089
[8]    train-mlogloss:0.830102 test-mlogloss:1.55741
[9]    train-mlogloss:0.769675 test-mlogloss:1.53544
[10]   train-mlogloss:0.714516 test-mlogloss:1.52551
[11]   train-mlogloss:0.665267 test-mlogloss:1.51696
[12]   train-mlogloss:0.619384 test-mlogloss:1.49978
[13]   train-mlogloss:0.576274 test-mlogloss:1.4831

```

```
[14] train-mlogloss:0.536544 test-mlogloss:1.47876
[15] train-mlogloss:0.500695 test-mlogloss:1.46847
[16] train-mlogloss:0.467632 test-mlogloss:1.45935
[17] train-mlogloss:0.437541 test-mlogloss:1.45583
[18] train-mlogloss:0.408675 test-mlogloss:1.44692
[19] train-mlogloss:0.383777 test-mlogloss:1.44072
[20] train-mlogloss:0.360373 test-mlogloss:1.4336
[21] train-mlogloss:0.338986 test-mlogloss:1.42599
[22] train-mlogloss:0.319027 test-mlogloss:1.42002
[23] train-mlogloss:0.300363 test-mlogloss:1.41875
[24] train-mlogloss:0.282776 test-mlogloss:1.41734
[25] train-mlogloss:0.266641 test-mlogloss:1.4144
[26] train-mlogloss:0.251735 test-mlogloss:1.4089
[27] train-mlogloss:0.23801 test-mlogloss:1.40467
[28] train-mlogloss:0.225228 test-mlogloss:1.39979
[29] train-mlogloss:0.213173 test-mlogloss:1.39832
[30] train-mlogloss:0.202329 test-mlogloss:1.39162
[31] train-mlogloss:0.192251 test-mlogloss:1.38982
[32] train-mlogloss:0.182745 test-mlogloss:1.38794
[33] train-mlogloss:0.173866 test-mlogloss:1.38432
[34] train-mlogloss:0.165626 test-mlogloss:1.38362
[35] train-mlogloss:0.157825 test-mlogloss:1.38073
[36] train-mlogloss:0.150726 test-mlogloss:1.37446
[37] train-mlogloss:0.144115 test-mlogloss:1.36934
[38] train-mlogloss:0.137944 test-mlogloss:1.36871
[39] train-mlogloss:0.131803 test-mlogloss:1.36586
[40] train-mlogloss:0.126562 test-mlogloss:1.36579
[41] train-mlogloss:0.121449 test-mlogloss:1.36793
[42] train-mlogloss:0.11663 test-mlogloss:1.36744
[43] train-mlogloss:0.112019 test-mlogloss:1.36535
[44] train-mlogloss:0.107669 test-mlogloss:1.36101
[45] train-mlogloss:0.103655 test-mlogloss:1.36156
[46] train-mlogloss:0.0998826 test-mlogloss:1.36127
[47] train-mlogloss:0.0963644 test-mlogloss:1.36075
[48] train-mlogloss:0.0931038 test-mlogloss:1.36163
[49] train-mlogloss:0.089967 test-mlogloss:1.36219
[50] train-mlogloss:0.0868146 test-mlogloss:1.36203
Stopping. Best iteration:
[47] train-mlogloss:0.0963644+0.00289686 test-mlogloss:1.3607
5+0.0750034

(173, 17699) (173,)
```

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
recated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
recated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/sklearn/model_selection/_split.py:652: Warning: The l
east populated class in y has only 1 members, which is too few. The
minimum number of members in any class cannot be less than n_splits=
5.
    % (min_groups, self.n_splits)), Warning)
```



```
[0]      train-mlogloss:1.63223  test-mlogloss:1.75837
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

```
Will train until test-mlogloss hasn't improved in 3 rounds.
```

```
[1]      train-mlogloss:1.49551  test-mlogloss:1.74583
[2]      train-mlogloss:1.36945  test-mlogloss:1.73475
[3]      train-mlogloss:1.25709  test-mlogloss:1.71828
[4]      train-mlogloss:1.16214  test-mlogloss:1.70871
[5]      train-mlogloss:1.07222  test-mlogloss:1.6938
[6]      train-mlogloss:0.991825 test-mlogloss:1.68509
[7]      train-mlogloss:0.918484 test-mlogloss:1.66968
[8]      train-mlogloss:0.851313 test-mlogloss:1.6629
[9]      train-mlogloss:0.792919 test-mlogloss:1.66007
[10]     train-mlogloss:0.739857 test-mlogloss:1.65137
[11]     train-mlogloss:0.689433 test-mlogloss:1.64202
[12]     train-mlogloss:0.641803 test-mlogloss:1.63549
[13]     train-mlogloss:0.598556 test-mlogloss:1.62549
[14]     train-mlogloss:0.560132 test-mlogloss:1.6211
[15]     train-mlogloss:0.524159 test-mlogloss:1.61161
[16]     train-mlogloss:0.490788 test-mlogloss:1.60567
[17]     train-mlogloss:0.459331 test-mlogloss:1.60326
[18]     train-mlogloss:0.429979 test-mlogloss:1.60223
[19]     train-mlogloss:0.403761 test-mlogloss:1.59955
[20]     train-mlogloss:0.37965  test-mlogloss:1.59806
[21]     train-mlogloss:0.35745  test-mlogloss:1.59249
[22]     train-mlogloss:0.337003 test-mlogloss:1.58813
[23]     train-mlogloss:0.317299 test-mlogloss:1.58244
[24]     train-mlogloss:0.298962 test-mlogloss:1.58186
[25]     train-mlogloss:0.282227 test-mlogloss:1.58003
[26]     train-mlogloss:0.266448 test-mlogloss:1.5794
[27]     train-mlogloss:0.251349 test-mlogloss:1.57943
[28]     train-mlogloss:0.238281 test-mlogloss:1.57595
[29]     train-mlogloss:0.226297 test-mlogloss:1.57236
[30]     train-mlogloss:0.214872 test-mlogloss:1.56935
[31]     train-mlogloss:0.204264 test-mlogloss:1.57453
[32]     train-mlogloss:0.194086 test-mlogloss:1.57158
[33]     train-mlogloss:0.184555 test-mlogloss:1.57006
```

```
Stopping. Best iteration:
```

```
[30]      train-mlogloss:0.214872+0.00619151      test-mlogloss:1.5693
5+0.0511381
```

```
(182, 18391) (182,)
```

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.62447  test-mlogloss:1.75387
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.48255  test-mlogloss:1.71667
[2]      train-mlogloss:1.35756  test-mlogloss:1.69529
[3]      train-mlogloss:1.24316  test-mlogloss:1.67548
[4]      train-mlogloss:1.14576  test-mlogloss:1.65709
[5]      train-mlogloss:1.05567  test-mlogloss:1.63905
[6]      train-mlogloss:0.978286 test-mlogloss:1.62344
[7]      train-mlogloss:0.905318 test-mlogloss:1.61347
[8]      train-mlogloss:0.838969 test-mlogloss:1.59276
[9]      train-mlogloss:0.780397 test-mlogloss:1.57982
[10]     train-mlogloss:0.725173 test-mlogloss:1.5671
[11]     train-mlogloss:0.675283 test-mlogloss:1.55172
[12]     train-mlogloss:0.629815 test-mlogloss:1.53812
[13]     train-mlogloss:0.58697  test-mlogloss:1.52861
[14]     train-mlogloss:0.547426 test-mlogloss:1.52199
[15]     train-mlogloss:0.511055 test-mlogloss:1.51063
[16]     train-mlogloss:0.478492 test-mlogloss:1.49976
[17]     train-mlogloss:0.449152 test-mlogloss:1.49007
[18]     train-mlogloss:0.420684 test-mlogloss:1.48316
[19]     train-mlogloss:0.394375 test-mlogloss:1.48049
[20]     train-mlogloss:0.369547 test-mlogloss:1.47542
[21]     train-mlogloss:0.347945 test-mlogloss:1.46721
[22]     train-mlogloss:0.3266    test-mlogloss:1.46134
[23]     train-mlogloss:0.30764  test-mlogloss:1.45787
[24]     train-mlogloss:0.289906 test-mlogloss:1.45172
[25]     train-mlogloss:0.273812 test-mlogloss:1.44993
[26]     train-mlogloss:0.258596 test-mlogloss:1.44619
[27]     train-mlogloss:0.244081 test-mlogloss:1.44502
[28]     train-mlogloss:0.231382 test-mlogloss:1.44122
[29]     train-mlogloss:0.219071 test-mlogloss:1.4406
[30]     train-mlogloss:0.207985 test-mlogloss:1.4393
[31]     train-mlogloss:0.197485 test-mlogloss:1.43418
[32]     train-mlogloss:0.188008 test-mlogloss:1.43474
[33]     train-mlogloss:0.17868  test-mlogloss:1.4357
[34]     train-mlogloss:0.169898 test-mlogloss:1.43564
```

Stopping. Best iteration:

```
[31]      train-mlogloss:0.197485+0.00356478      test-mlogloss:1.4341
8+0.0646309
```

```
(181, 21107) (181,)
```

```
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:587: FutureWarning: Series.base is de
precated and will be removed in a future version
    if getattr(data, 'base', None) is not None and \
/home/imlay/programs.installed/anaconda3/envs/cs418env/lib/python3.7
/site-packages/xgboost/core.py:588: FutureWarning: Series.base is de
precated and will be removed in a future version
    data.base is not None and isinstance(data, np.ndarray) \
```

```
[0]      train-mlogloss:1.6181    test-mlogloss:1.74395
Multiple eval metrics have been passed: 'test-mlogloss' will be used
for early stopping.
```

Will train until test-mlogloss hasn't improved in 3 rounds.

```
[1]      train-mlogloss:1.47267    test-mlogloss:1.70457
[2]      train-mlogloss:1.34405    test-mlogloss:1.67228
[3]      train-mlogloss:1.22628    test-mlogloss:1.64019
[4]      train-mlogloss:1.12568    test-mlogloss:1.61042
[5]      train-mlogloss:1.03772    test-mlogloss:1.58886
[6]      train-mlogloss:0.957709   test-mlogloss:1.57296
[7]      train-mlogloss:0.884824   test-mlogloss:1.56142
[8]      train-mlogloss:0.819988   test-mlogloss:1.54574
[9]      train-mlogloss:0.760515   test-mlogloss:1.53351
[10]     train-mlogloss:0.706686   test-mlogloss:1.51831
[11]     train-mlogloss:0.659158   test-mlogloss:1.50782
[12]     train-mlogloss:0.614127   test-mlogloss:1.50188
[13]     train-mlogloss:0.574015   test-mlogloss:1.49195
[14]     train-mlogloss:0.536618   test-mlogloss:1.4839
[15]     train-mlogloss:0.503651   test-mlogloss:1.47697
[16]     train-mlogloss:0.471164   test-mlogloss:1.46727
[17]     train-mlogloss:0.440918   test-mlogloss:1.46412
[18]     train-mlogloss:0.412768   test-mlogloss:1.45738
[19]     train-mlogloss:0.387996   test-mlogloss:1.45539
[20]     train-mlogloss:0.364644   test-mlogloss:1.4514
[21]     train-mlogloss:0.342551   test-mlogloss:1.45009
[22]     train-mlogloss:0.321935   test-mlogloss:1.44562
[23]     train-mlogloss:0.303158   test-mlogloss:1.44178
[24]     train-mlogloss:0.285473   test-mlogloss:1.43861
[25]     train-mlogloss:0.268985   test-mlogloss:1.43143
[26]     train-mlogloss:0.254174   test-mlogloss:1.42892
[27]     train-mlogloss:0.239677   test-mlogloss:1.43005
[28]     train-mlogloss:0.226742   test-mlogloss:1.42529
[29]     train-mlogloss:0.21438    test-mlogloss:1.42214
[30]     train-mlogloss:0.20291    test-mlogloss:1.42029
[31]     train-mlogloss:0.192466   test-mlogloss:1.41851
[32]     train-mlogloss:0.182558   test-mlogloss:1.41596
[33]     train-mlogloss:0.173124   test-mlogloss:1.41302
[34]     train-mlogloss:0.164276   test-mlogloss:1.41148
[35]     train-mlogloss:0.156177   test-mlogloss:1.41022
[36]     train-mlogloss:0.148588   test-mlogloss:1.41047
[37]     train-mlogloss:0.141523   test-mlogloss:1.41103
[38]     train-mlogloss:0.135018   test-mlogloss:1.41255
```

Stopping. Best iteration:

```
[35]      train-mlogloss:0.156177+0.00442193    test-mlogloss:1.4102
2+0.0546241
```

```
(312, 18172) (312,)
```

```

-----
-----
NameError                                Traceback (most recent call
last)
<ipython-input-12-19c89eae0926> in <module>
      1 algs,results=mainLoop(df,data_dir,tissue_dir)
----> 2 print(d)

NameError: name 'd' is not defined

```

Model and Score Saving

```

In [28]: from sklearn.externals import joblib
results.to_csv("models/results.tsv",sep="\t",header=['Score'])
[joblib.dump(algs[x], 'models/{}_model.pkl'.format(results.index.values
[x])) for x in range(len(results))]

```

```

Out[28]: [['models/Adipose Tissue_model.pkl'],
['models/Blood Vessel_model.pkl'],
['models/Brain_model.pkl'],
['models/Breast_model.pkl'],
['models/Blood_model.pkl'],
['models/Skin_model.pkl'],
['models/Colon_model.pkl'],
['models/Esophagus_model.pkl'],
['models/Heart_model.pkl'],
['models/Lung_model.pkl'],
['models/Muscle_model.pkl'],
['models/Nerve_model.pkl'],
['models/Pancreas_model.pkl'],
['models/Stomach_model.pkl'],
['models/Testis_model.pkl'],
['models/Thyroid_model.pkl']]

```

Model and Score Loading

```
In [29]: results2=pd.read_csv("models/results.tsv",sep="\t",index_col=0)
results2
```

Out[29]:

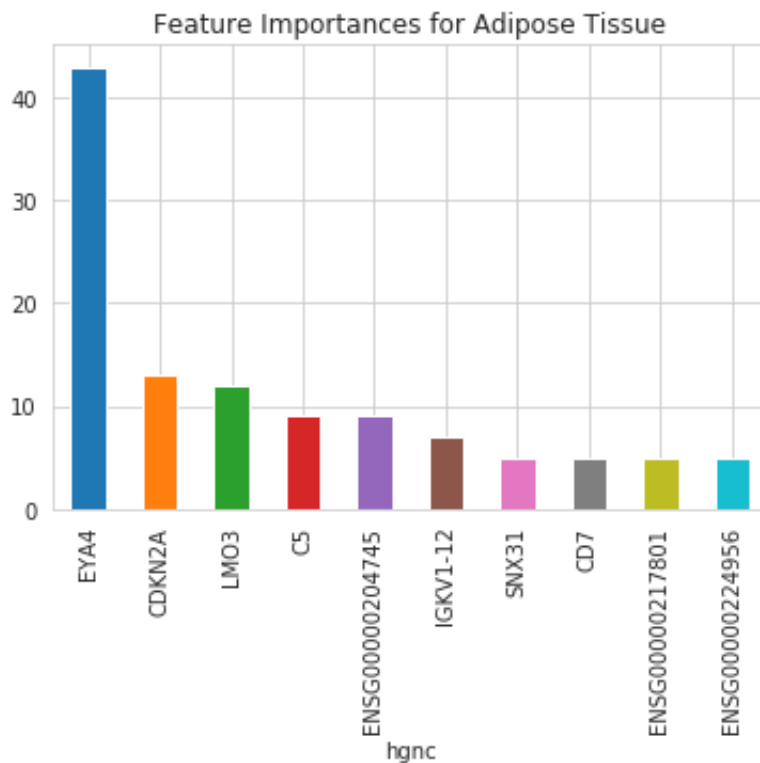
	Score
Adipose Tissue	0.337500
Blood Vessel	0.448905
Brain	0.530000
Breast	0.275862
Blood	0.314815
Skin	0.373961
Colon	0.450980
Esophagus	0.384365
Heart	0.488889
Lung	0.348837
Muscle	0.411765
Nerve	0.472000
Pancreas	0.346667
Stomach	0.316456
Testis	0.320513
Thyroid	0.417910

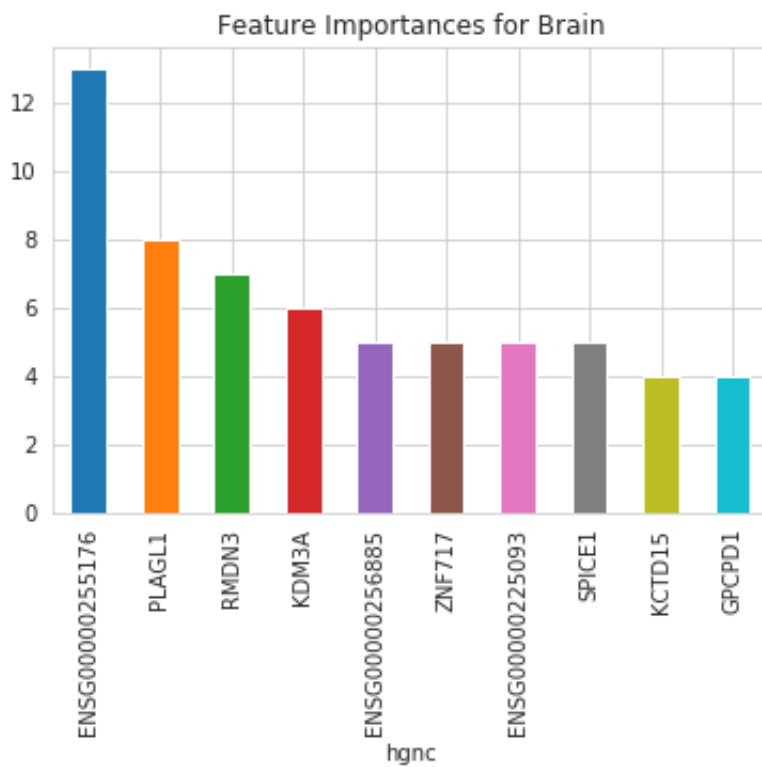
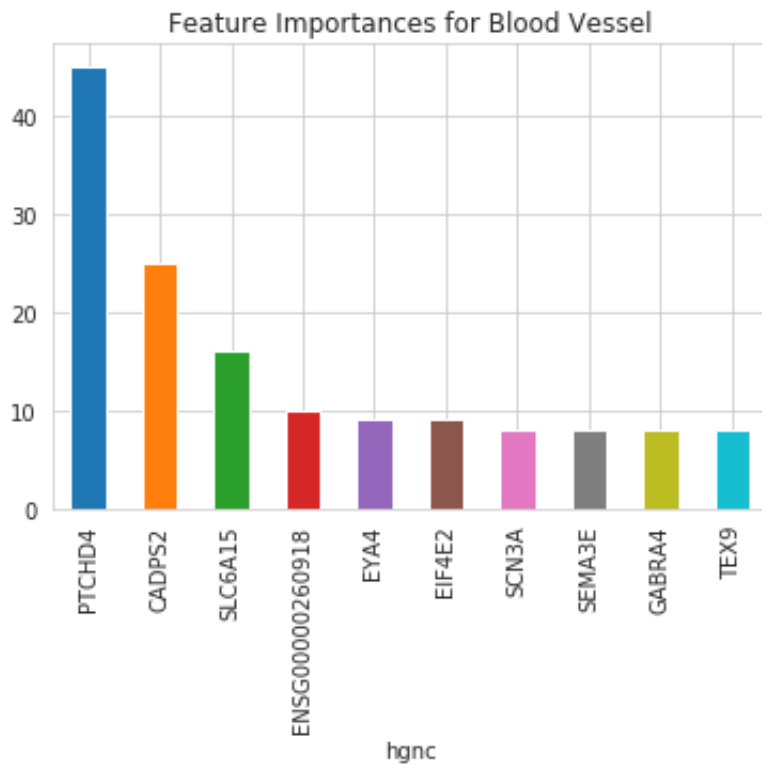
```
In [ ]: algs2=[joblib.load('models/{_}model.pkl'.format(results.index.values[x
])) for x in range(len(results))]
```

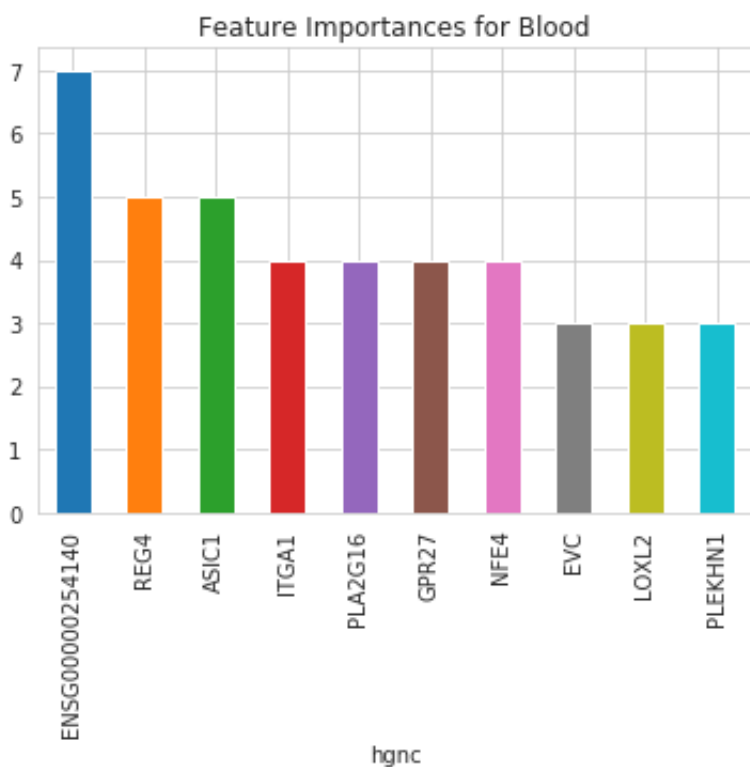
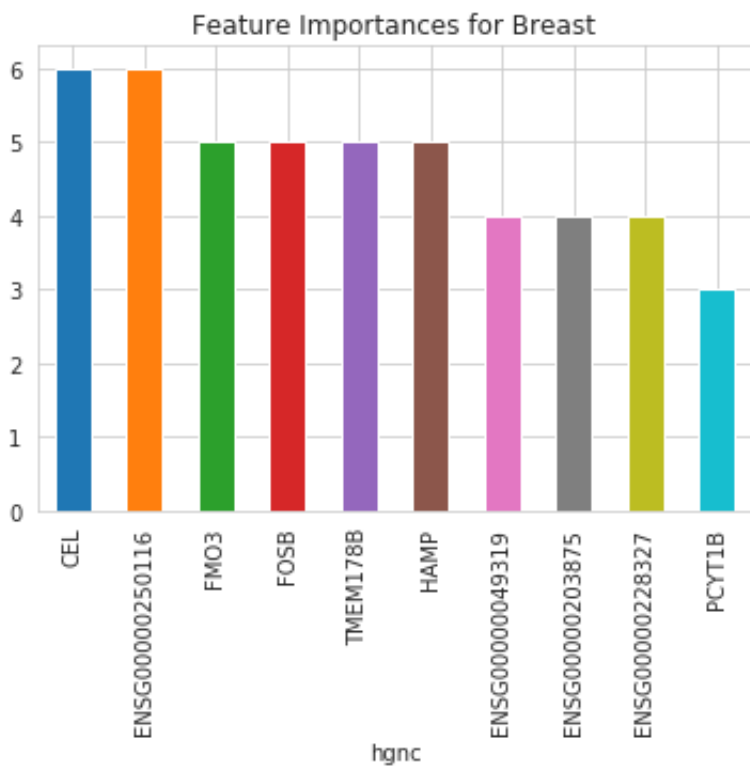
Model Evaluation

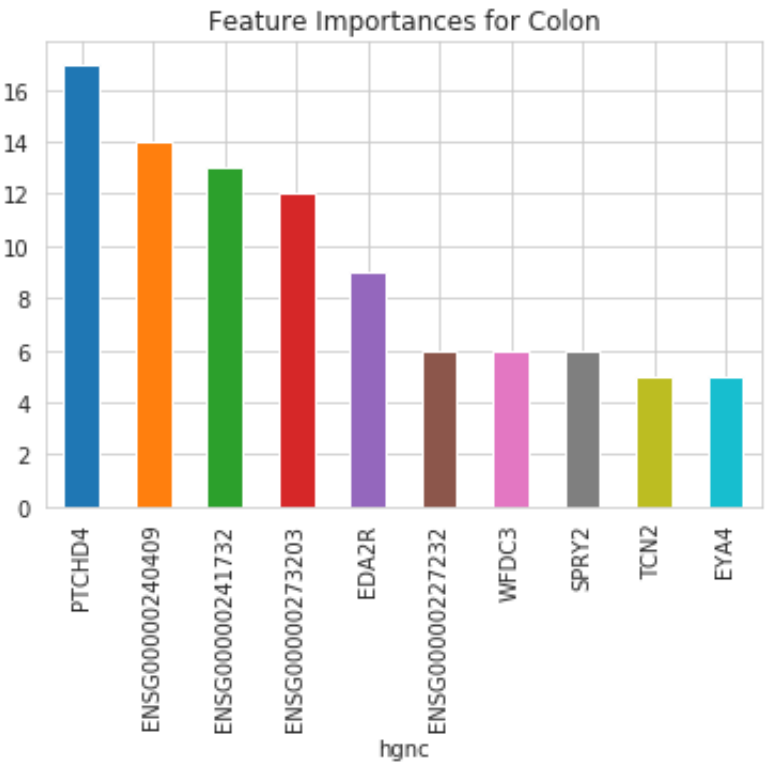
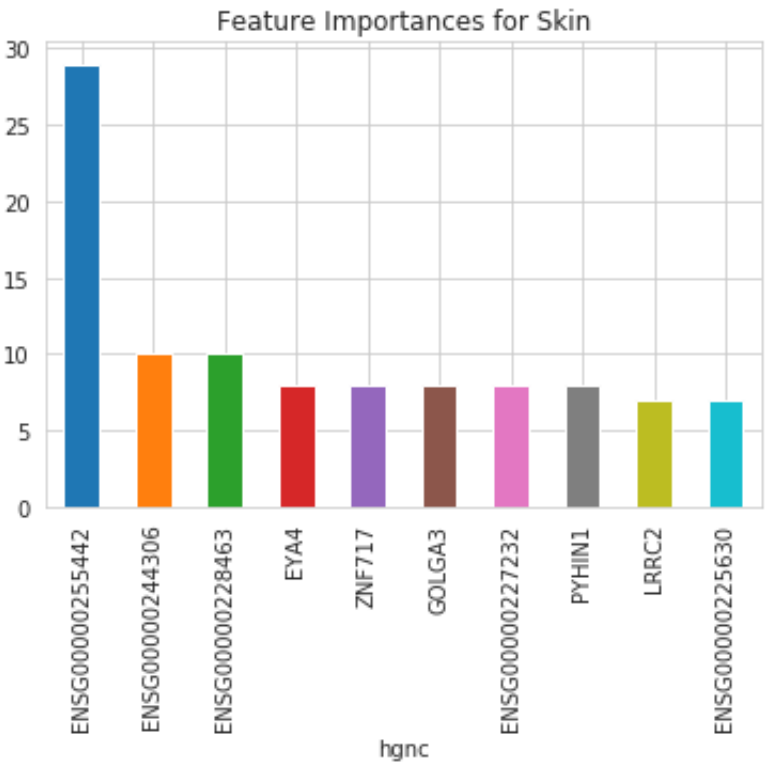
```
In [ ]: uniprot=pd.read_csv("annotation/uniprot_meta.tsv",sep="\t",names=["Uni
prot","HGNC_Species","Full HGNC","Location","Type","Desc"])
biomart=pd.read_csv("annotation/2018-04-12_biomart_ensemblGene_hgnc_un
iprot.tab",sep="\t")
sel=pd.isnull(biomart['hgnc'])
biomart.loc[sel,"hgnc"]=biomart.loc[sel,"ensembl"] # fills in missing
values of hgnc with ensembl
biomart = biomart.drop_duplicates(subset='ensembl', keep='first')
biomart=biomart.set_index(biomart['ensembl'])
```

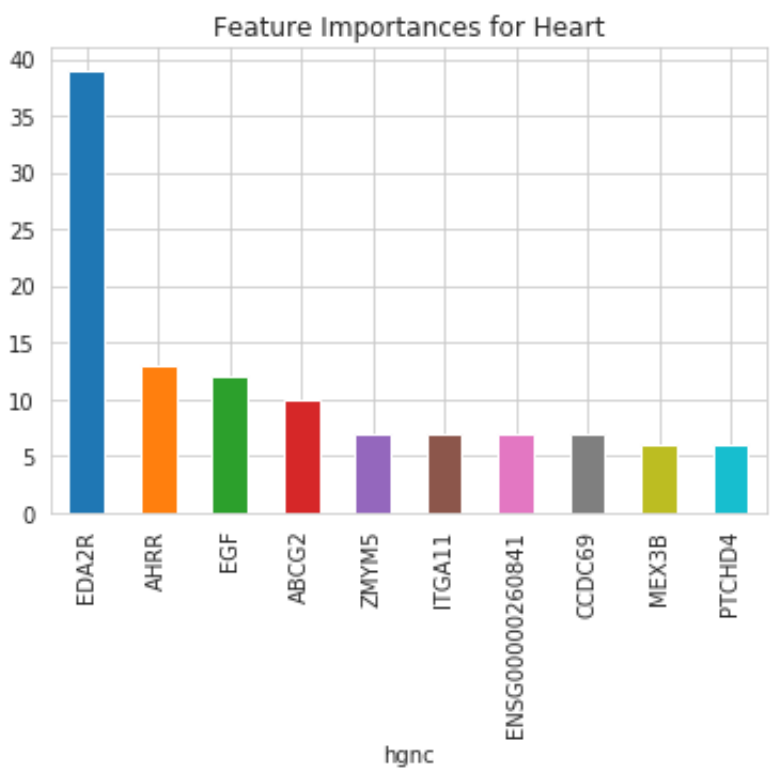
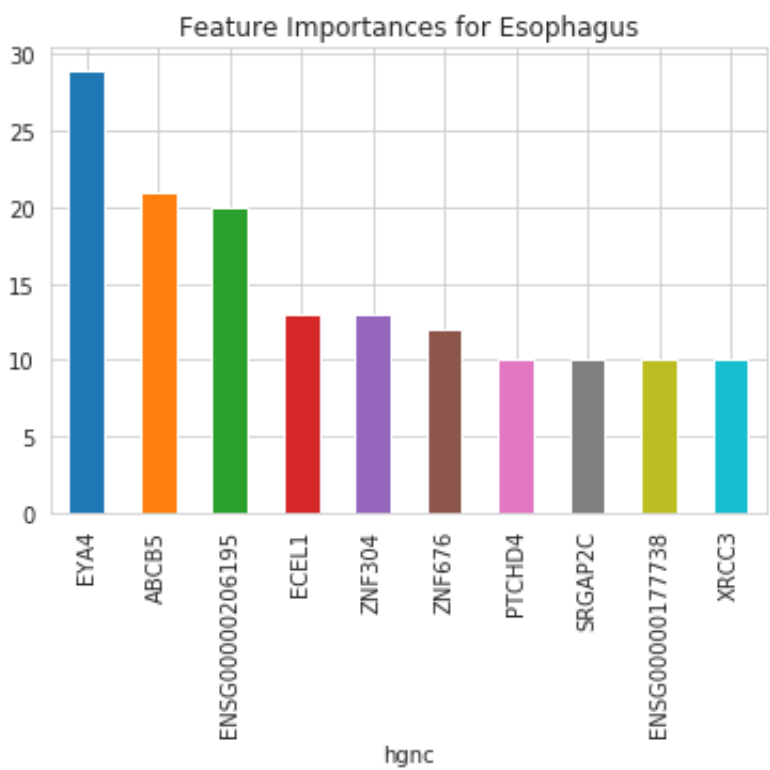
```
In [82]: from xgboost import plot_tree
res=[]
for x in range(len(results.index.values)):
    alg=algs[x]
    TISSUE=results.index.values[x]
    feat_imp = pd.Series(alg.get_booster().get_score(importance_type='
weight')).sort_values(ascending=False)[:200]
    feat_imp=feat_imp.to_frame('Score').join(biomart)
    sel=pd.isnull(feat_imp['hgnc'])
    feat_imp.loc[sel,"hgnc"]=feat_imp.index.values[sel]
    feat_imp=feat_imp.set_index(feat_imp['hgnc'])
    x=feat_imp['Score'][:10].plot(kind='bar', title='Feature Importanc
es for {}'.format(TISSUE))
    #x.set_ylabel('Feature Importance Score')
    plt.show(x)
    plt.savefig(fname='plots/{}_feat_imp.png'.format(TISSUE))
```

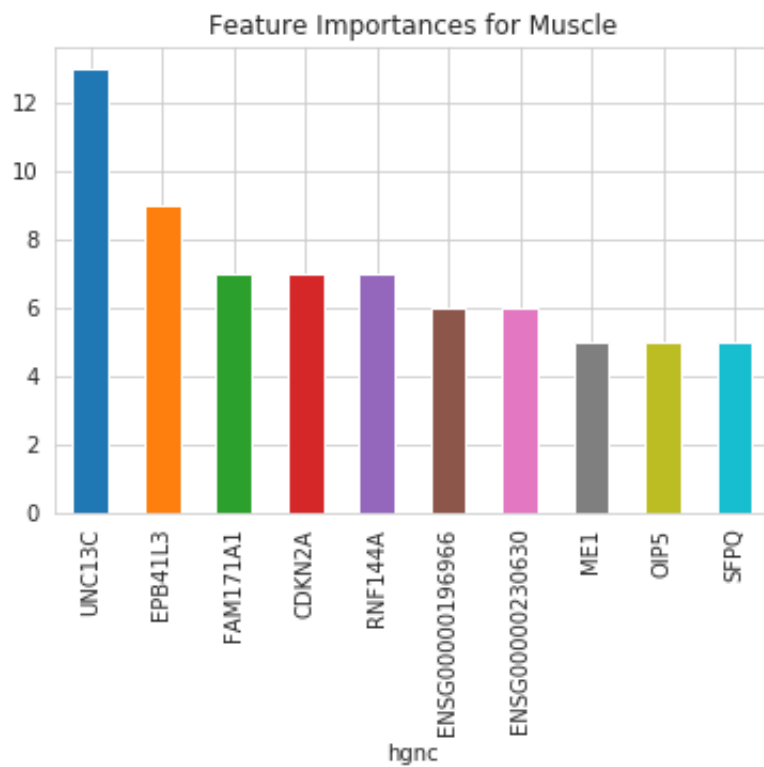
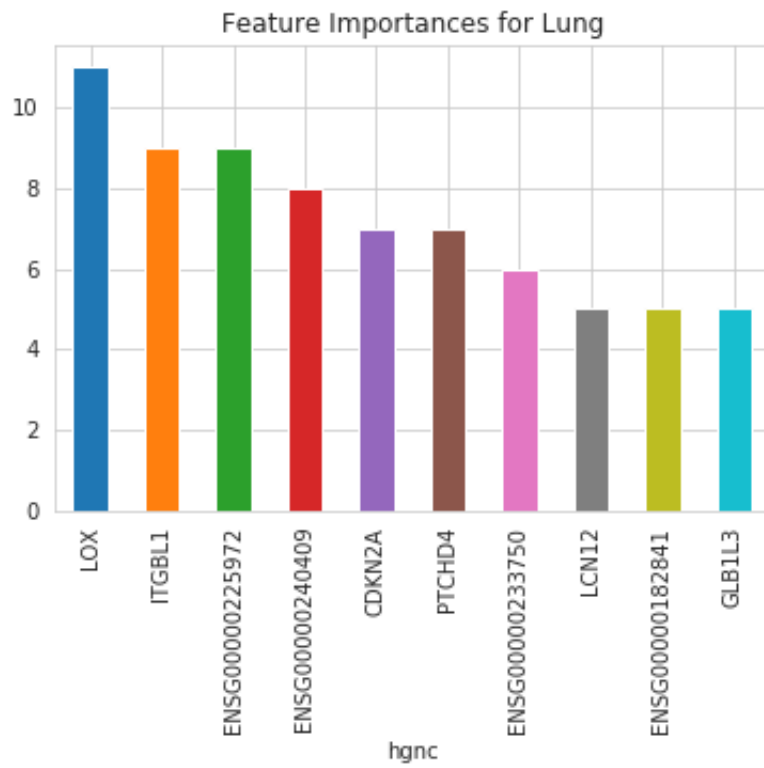


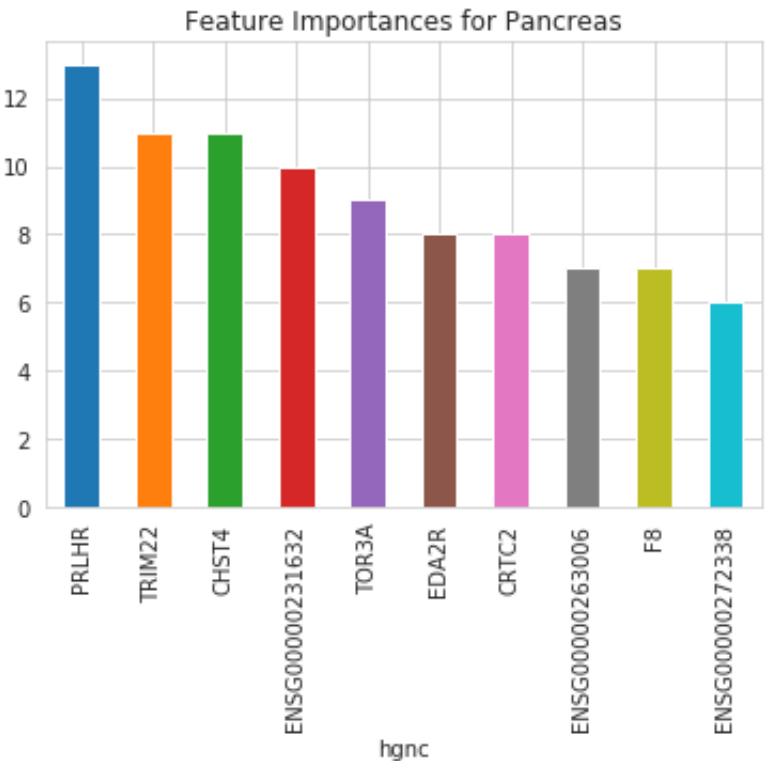
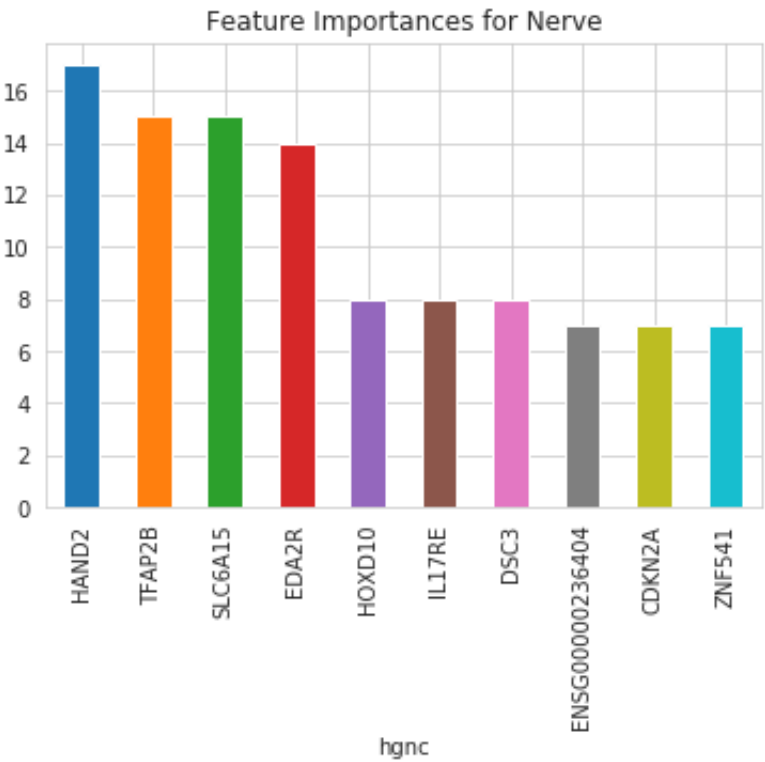


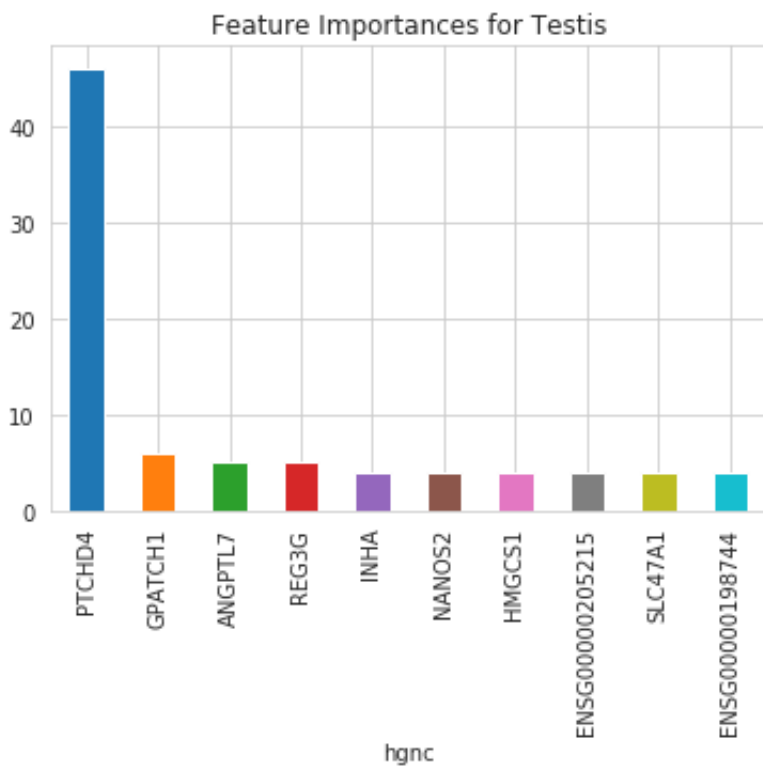
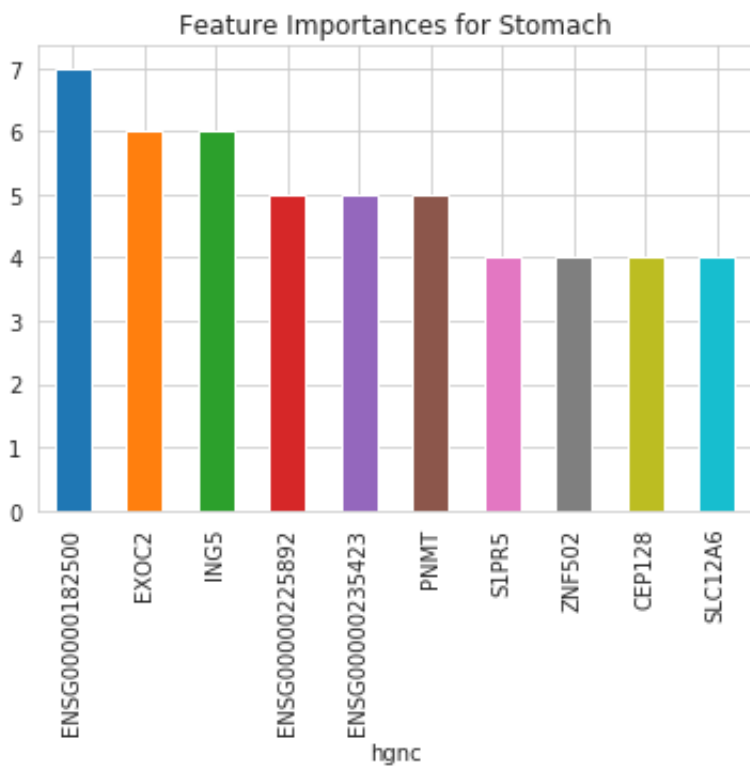


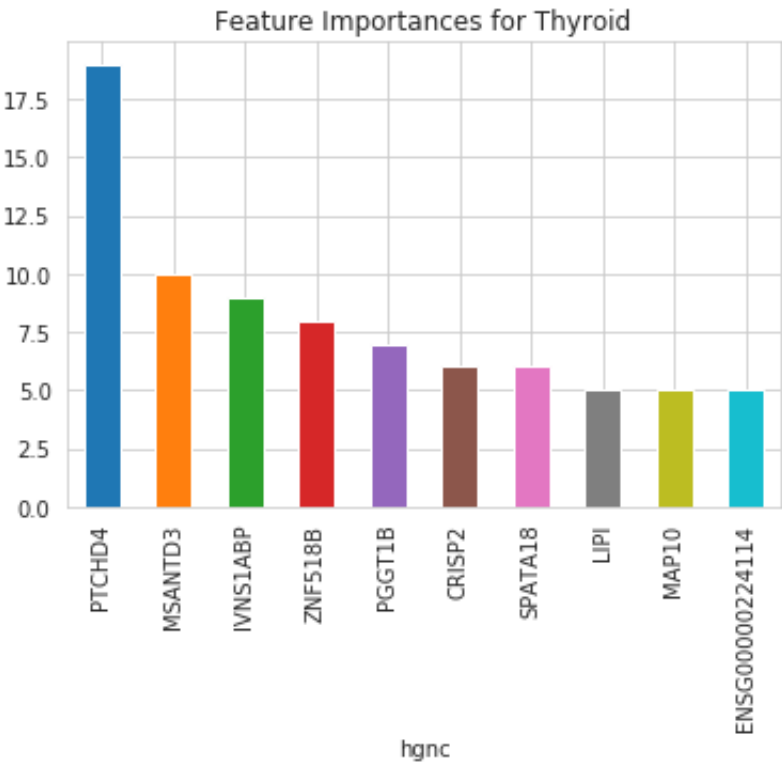






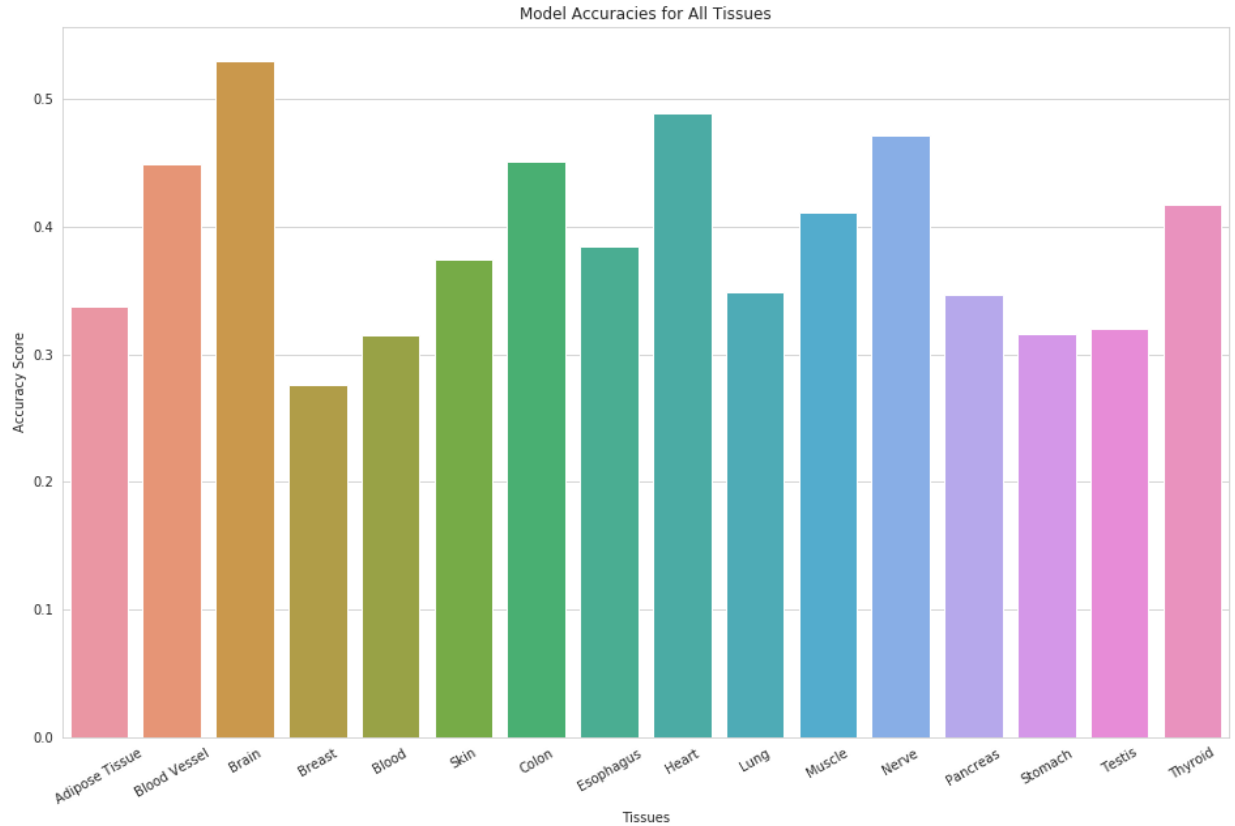






<Figure size 432x288 with 0 Axes>


```
In [40]: plt.figure(figsize=(16, 10))
ax=sns.barplot(x=results.index.values,y=results)
x=ax.set_title("Model Accuracies for All Tissues")
x=ax.set_xlabel("Tissues")
x=ax.set_ylabel("Accuracy Score")
x=ax.set_xticklabels(labels=results.index.values,rotation=30)
```



```
In [ ]:
```