ex:4
Syed Mohsin Ali

Question 1:

k = 1
Distance Metric = Euclidean
Preprocessing = MinMax

avg. score for set1 : 0.854652014652
avg. score for set2 : 0.931691221608
avg. score for set3 : 0.737569252078
avg. score for set4 : 0.847986924077
avg. score for set5 : 0.859010989011

Question 2:

Best Params Set1:

k = 3
preprocessor = Scalar
distance method = **Euclidean**

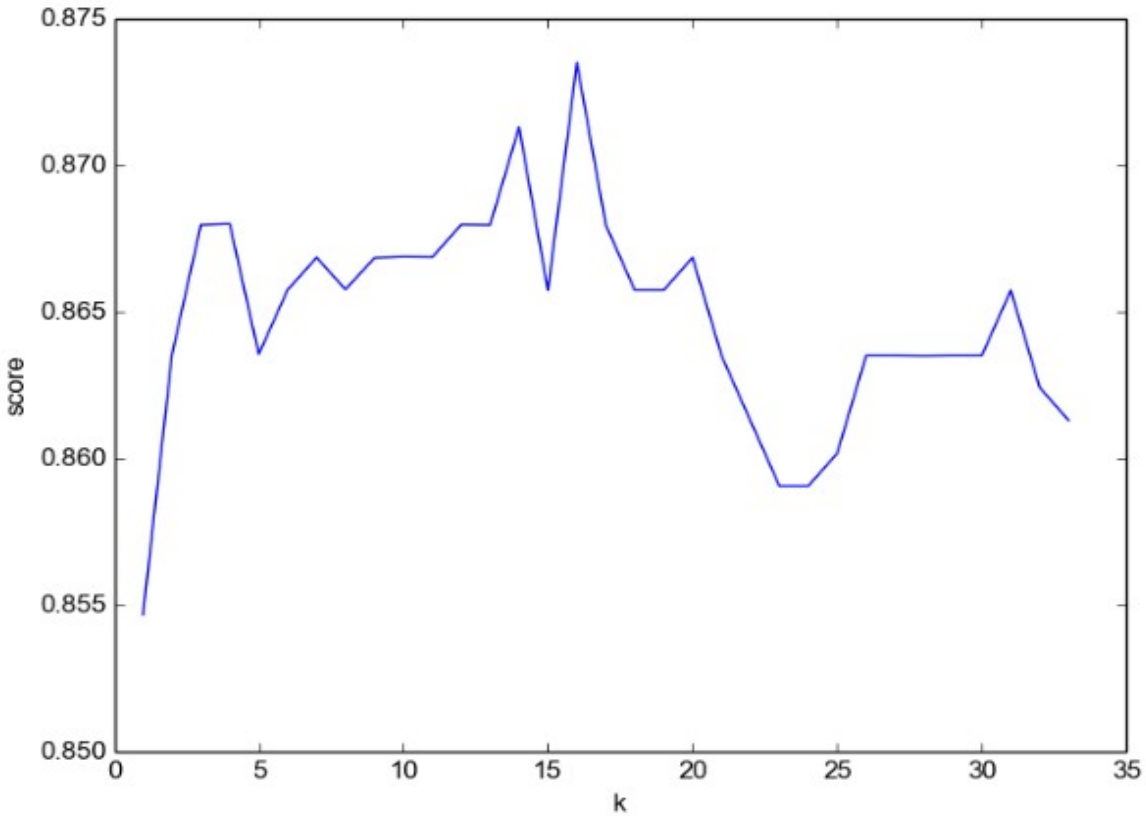
score = 0.963348680172

Best Params Set2:

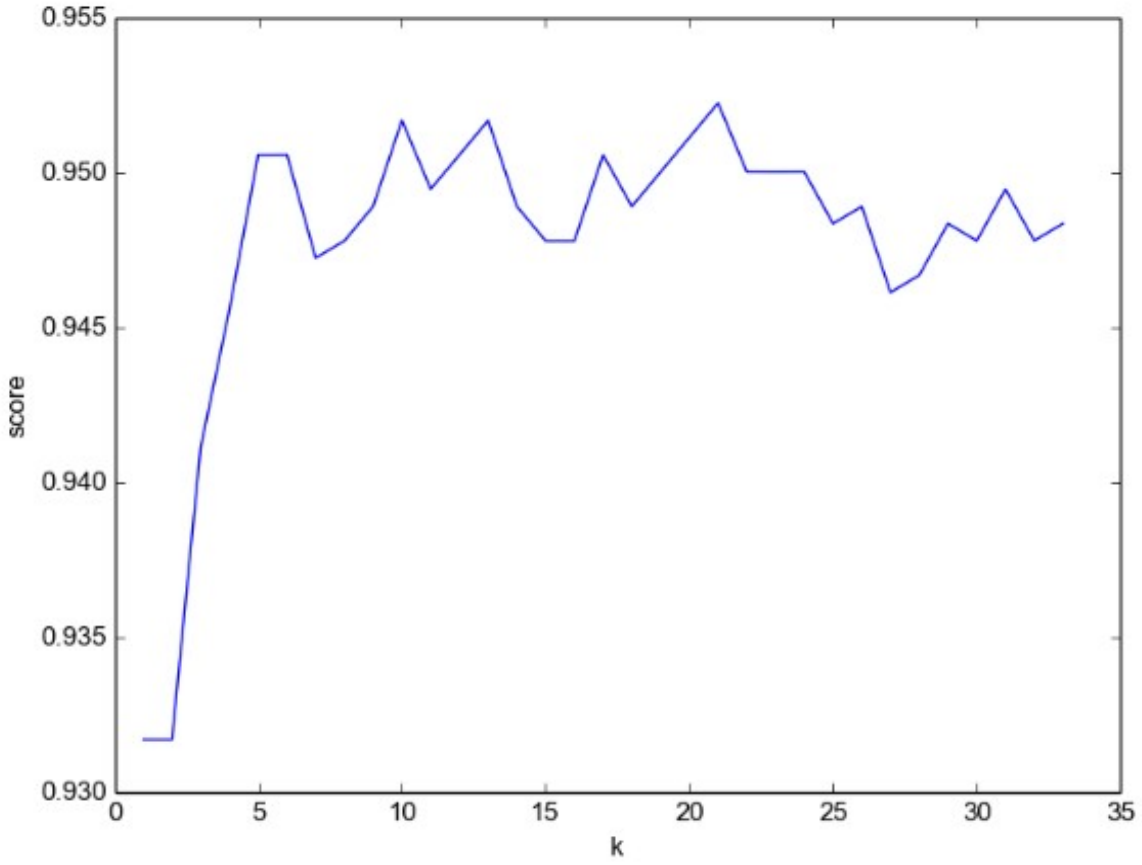
k = 5
preprocessor = Scalar
distance method = **Manhattan**

score = 0.967237569061

Set1: MinMax, Euclidean



Set2: MinMax, Euclidean



Question 3:

```
[(0.520000, ParamSklearnClassifier(configuration={
    'balancing:strategy': 'weighting',
    'classifier:__choice__': 'extra_trees',
    'classifier:extra_trees:bootstrap': 'False',
    'classifier:extra_trees:criterion': 'entropy',
    'classifier:extra_trees:max_depth': 'None',
    'classifier:extra_trees:max_features': 1.71429054733,
    'classifier:extra_trees:min_samples_leaf': 1,
    'classifier:extra_trees:min_samples_split': 6,
    'classifier:extra_trees:min_weight_fraction_leaf': 0.0,
    'classifier:extra_trees:n_estimators': 100,
    'imputation:strategy': 'median',
    'one_hot_encoding:minimum_fraction': 0.00297257226834,
    'one_hot_encoding:use_minimum_fraction': 'True',
    'preprocessor:__choice__': 'polynomial',
    'preprocessor:polynomial:degree': 2,
    'preprocessor:polynomial:include_bias': 'False',
    'preprocessor:polynomial:interaction_only': 'True',
    'rescaling:__choice__': 'min/max'})),
  (0.480000, ParamSklearnClassifier(configuration={
    'balancing:strategy': 'none',
    'classifier:__choice__': 'qda',
    'classifier:qda:reg_param': 9.28492834134,
    'imputation:strategy': 'most_frequent',
    'one_hot_encoding:minimum_fraction': 0.159731683478,
    'one_hot_encoding:use_minimum_fraction': 'True',
    'preprocessor:__choice__': 'kitchen_sinks',
    'preprocessor:kitchen_sinks:gamma': 1.05043840055,
    'preprocessor:kitchen_sinks:n_components': 235,
    'rescaling:__choice__': 'min/max'})),
  ]
```

Question 4:

Data looked very simple so I used simple SVM. With bit of random parameter tuning I selected gamma 0.69 and kernel = rbf. As preprocessing I only used SCALAR.

FeedBACK

question1:
-about 3 hours
-i didnt had experience with numpy so it was good intro exercise for that
-i think exercise in implementing neural network would have been more fun
-compared to one of my friend im getting very good scores (in range 0.85-0.96) so im in bit of doubt if there might be some problem in my implementation. if rough result for one of dataset was provided then i could be sure.

question2:
-about 3 hours (time was mostly spent in waiting for results and making small changes to get them

in right format)
-i learned how frustating can it be to tune params especially if size of datasets start growing
-it was nice exercise, i wouldnt change anything

question3:
-about 3 hours
-all time was spent in downloading, starting and running vm
-i didnt fully understand the output of classifier.show_models() so i cant say i learned much
-i also wonder if one hour was enough for autosklearn. apparently results i got in it (0.52 and 0.48) are worse than ones i get with simple knn

question4:
-about 2 hours
-very open ended
-i liked the question but in the end I lacked the time to properly do it