BT5152

AY2018/2019 Semester 1 Week **11**

W11: Support Vector Machine

Tutorial Logistics

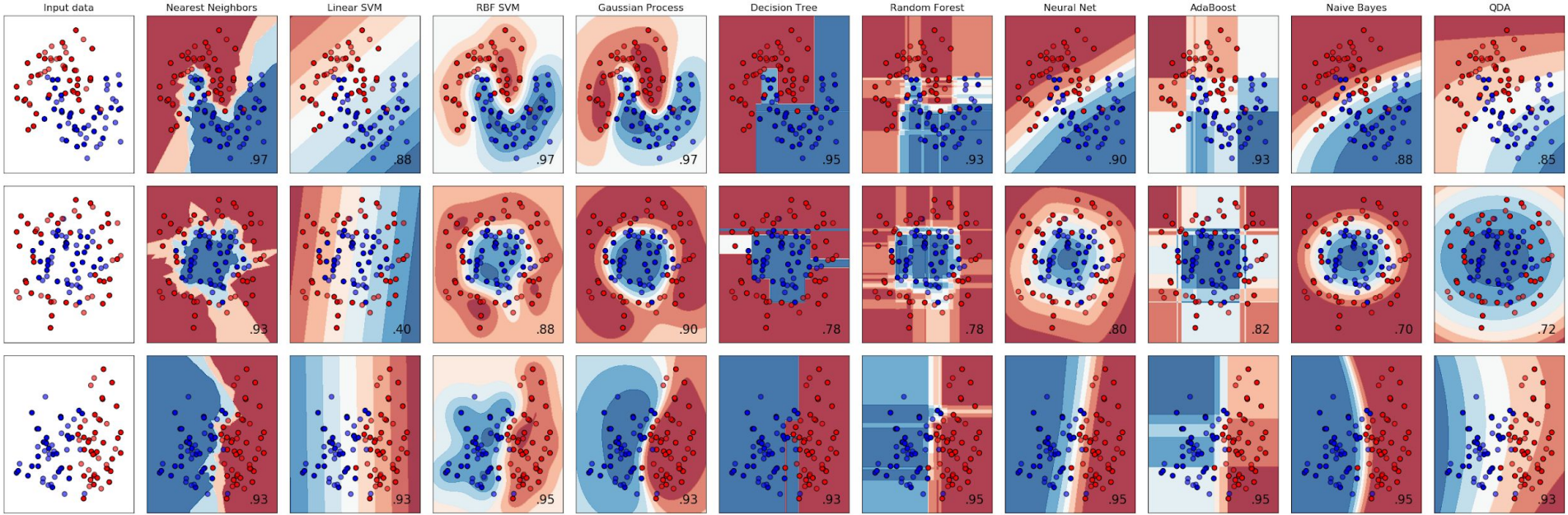
Swirl lesson on SVM is available on:

- https://github.com/weilu/BT5152
- https://github.com/kylase/BT5152

Things we covering today

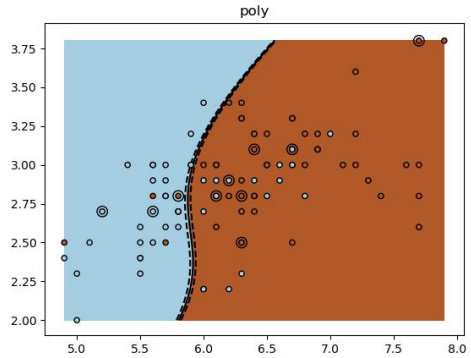
1. Brief talk on visualisation of ML
2. Training a linear SVM and visualise [Demo]
3. Training different SVM kernels and visualise them [10 to 20 mins]
4. Assignment 4 Reminders
5. Assignment 5 Question 1

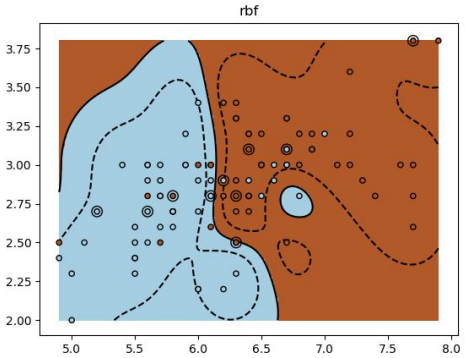
What different ML algorithm is doing (visually)?

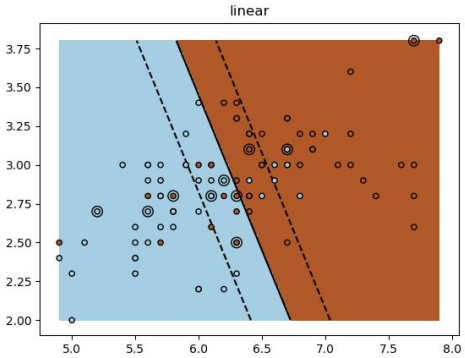


http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

Different SVM Kernels







http://scikit-learn.org/stable/auto_examples/exercises/plot_iris_exercise.html#sphx-glr-auto-examples-exercises-plot-iris-exercise-py

Sentiment Analysis (MovieLens)

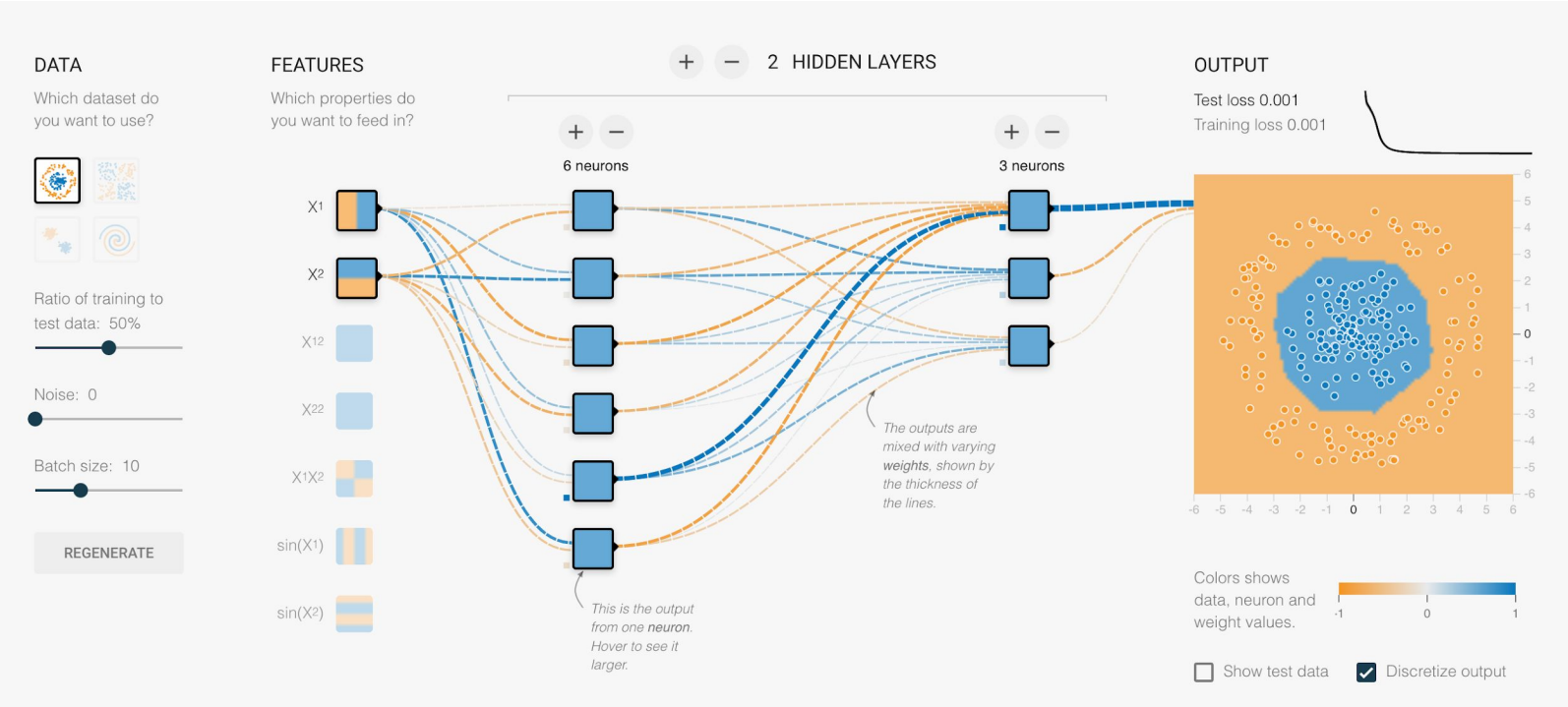
Naive Bayes

```
Confusion Matrix and Statistics

          Reference
Prediction   neg   pos
       neg  9001  2132
       pos  3499 10368

               Accuracy : 0.7748
                 95% CI : (0.7695, 0.7799)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5495
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7201
            Specificity : 0.8294
         Pos Pred Value : 0.8085
         Neg Pred Value : 0.7477
             Prevalence : 0.5000
         Detection Rate : 0.3600
   Detection Prevalence : 0.4453
      Balanced Accuracy : 0.7748
```

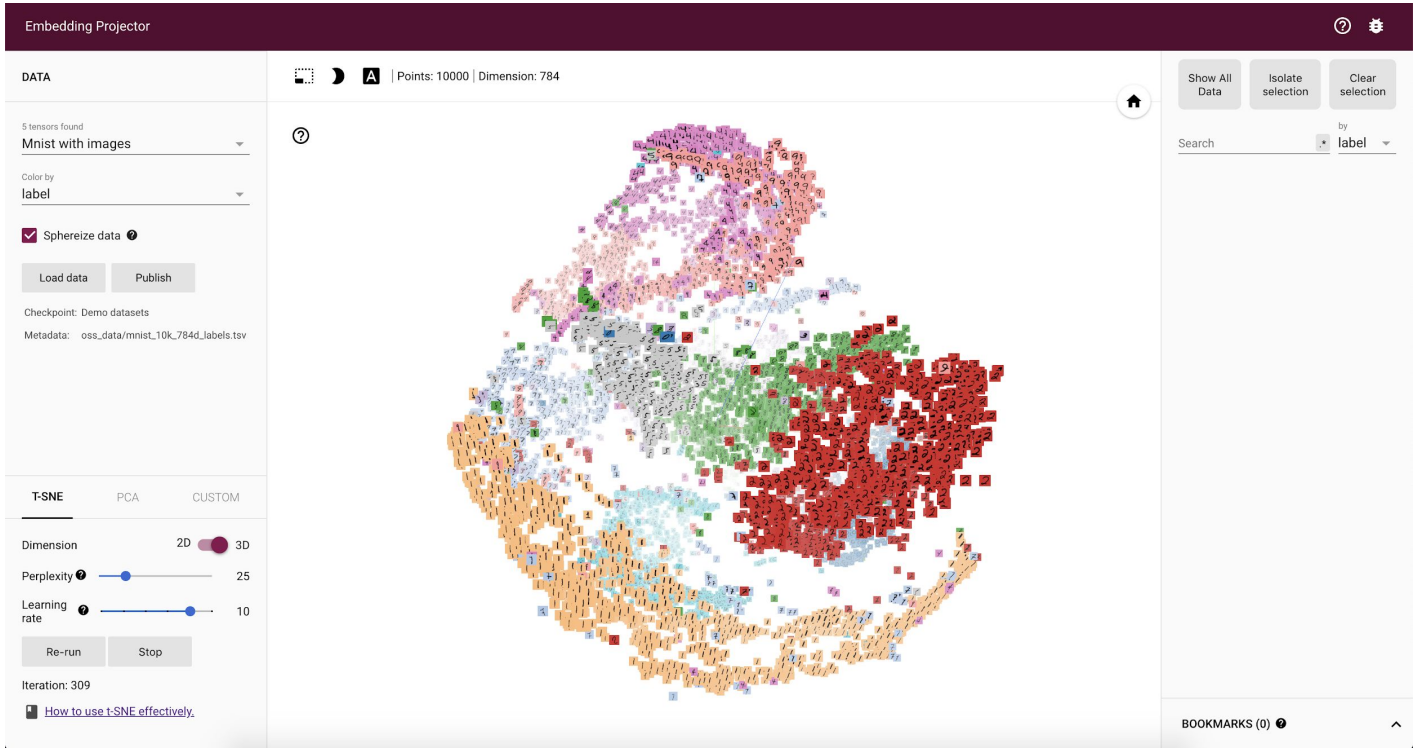
SVM (Linear Kernel)

```
Confusion Matrix and Statistics

          Reference
Prediction   neg   pos
       neg  9984  2075
       pos  2516 10425

               Accuracy : 0.8164
                 95% CI : (0.8115, 0.8211)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6327
 Mcnemar's Test P-Value : 8.37e-11

            Sensitivity : 0.7987
            Specificity : 0.8340
         Pos Pred Value : 0.8279
         Neg Pred Value : 0.8056
             Prevalence : 0.5000
         Detection Rate : 0.3994
   Detection Prevalence : 0.4824
      Balanced Accuracy : 0.8164
```

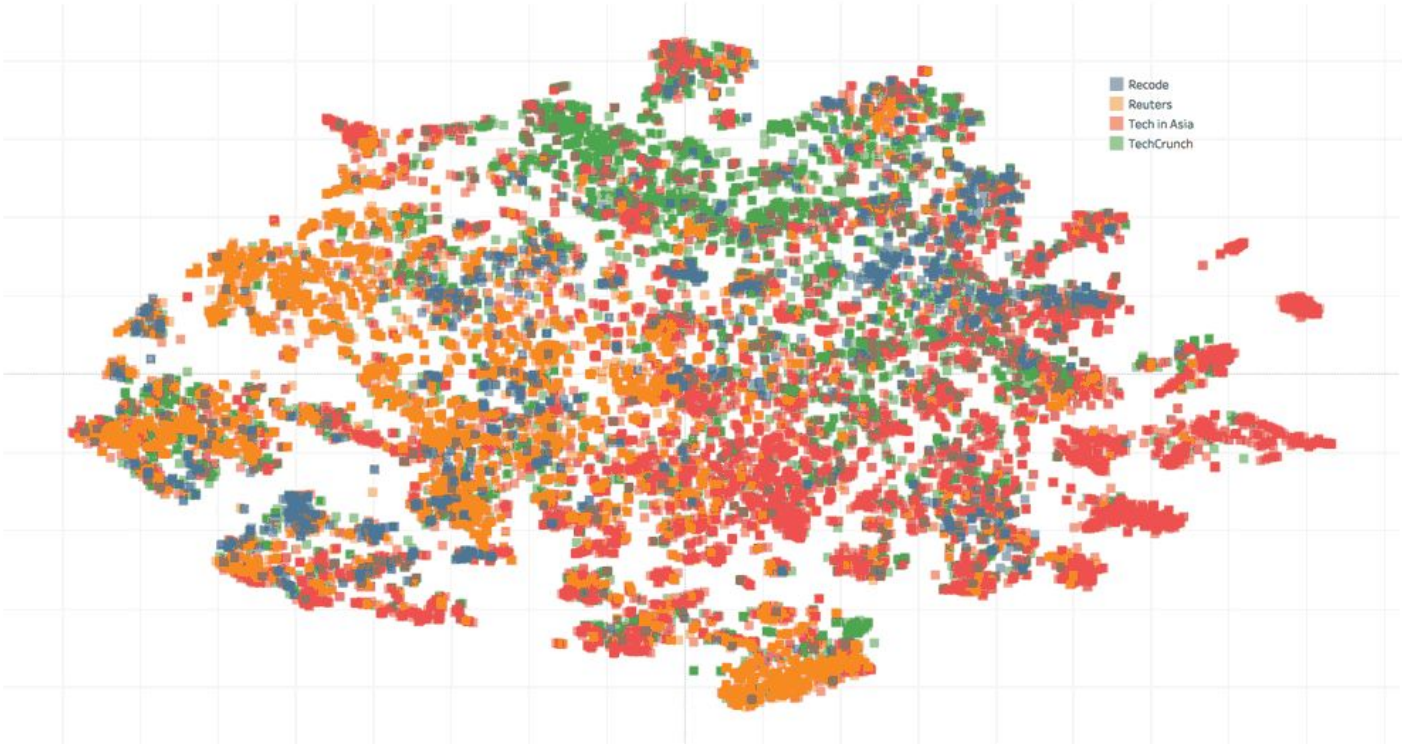
Tensorflow Playground



Tensorboard Projector



Visualising Topic Model (my previous work)



Training a SVM model using e1071

```
model <- svm(y ~ ., data=train, kernel=<kernel>, cost=<cost of misclassification>, <kernel_parameters>)
```

Visualising the kernel and prediction

```
plot(model, train)
```

Training a linear kernel

```
model.linear <- svm(y ~ ., data=train, kernel='linear', cost=10)
```

Dataset is available on https://github.com/kylase/BT5152/tree/master/BT5152_Tutorial_9

Assignment 4 Reminders

- Do not `setwd()`
- You can assume the data files are located in the **root directory** or **`data/ted-talk`** directory relative to where your **`code.R`** is
- If you are still stuck, please come to me after the class

Assignment 5 Question 1

Using `svm` on A3 dataset (`y ~ x4 + x8`) with different kernels and visualise it.

- Polynomial
- RBF
- Sigmoid