From linear classifiers to SVMs  
  
Ecole Centrale Nantes

Anna Hauschild, Jann Warnecke

SUPPORT VECTOR MACHINES

**a) Linear SVM**

• *Is it necessary to normalize the features in this case?*

Since the values of the data are between 0 and 1, it is already normalized. So, it is not necessary to normalize in this case.   
The reason that our dataset does not need normalization is because it is already normalized when we turned the coloured pictures into grey scale. In grey scale each pixel is ranged from 0 to 1, thus they are already normalized in a sense.

*• How are these probabilities computed?*

The probability model is created using cross validation, so the results can be slightly different than those obtained by predict. Also, it will produce meaningless results on very small datasets. Cross validation means that all data are at least once regarded as part of the 20% test set. One block at a time and then summarized as result at the end.

**b) Support vectors**

The model has 10000 parameters which are the pixels of the image. The hyperparameter C = 1 controls the trade-off between the slack variable penalty and the margin. The higher C the less error tolerance we have.

Support vectors are the datapoints near to the hyperplane, depending on the hyperparameter C. Here we see 74 support vectors. For displaying the images the data has to be reshaped into 2D.

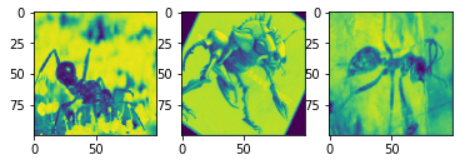


Figure 1 Display of some support vectors

**c) Kernel SVMs.**

The kernel is a mathematical function to manipulate the data. It transforms the data set, so that a nonlinear decision surface can be transformed into a linear equation in a higher number of dimension spaces.

When training an SVM with the **Radial Basis Function (RBF) kernel**, 81 support vectors are found and two hyperparameters must be considered: C and gamma. After the prediction 8 errors were found, the accurarcy is 0.65217.

When training with the **Polynomial function (poly) kernel**, 69 support vectors are found and four hyperparameters must be considered: gamma, degree, coef0 and C. After the prediction 7 errors were found, the accurarcy is 0.95652.

And when training with the **Sigmoid function (sigmoid) kernel**, also 69 support vectors are found and three hyperparameters must be considered: gamma, coef0 and C. After the prediction 8 errors were found, the accurarcy is 0.65217.

**d) Hyperparameter tuning**

Hyperparameters can't be learnt from the data. The performance improves with a more acceptable choice of hyperparameter tuning and selection techniques.

**The steps to tune the hyperparameters are**

1.Select the right type of model

2.Review the list of parameters of the model and build the hyperparameter space

3.Finding the methods for searching the hyperparameter space

4.Applying the cross-validation scheme approach

5.Assess the model score to evaluate the model

GridSearch helps us to search for the hyperparameter space (3.), applies the cross validation (4.) and asses the score (5.).

Using GridSearchCV the values C=10, gamma=0.001 and degree=2 are found as optimal hyperparameters.

2. PERFORMANCE MEASURES

**a) Performance report**

•RBF

TP = 11, TN =3, FP = 5, FN = 4 Total = 23  
Accuracy =0.61  
Precision = 0.69  
Recall = 0.73  
Specificity = 0.375  
F-measure = 0.71  
NVP = 0.43  
FPV = 0.31

Classification report

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Figure RBF Classification report

•Poly

TP = 15, TN = 0, FP = 8, FN = 0 Total = 23  
Accuracy =0.65  
Precision = 0.65  
Recall = 1  
Specificity = 0.0  
F-measure = 0.79  
NVP = nan  
FPV = 0.35

Classification report  
Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Figure 3 Poly Classification report

•Sigmoid

TP = 15, TN = 0, FP = 8, FN = 0 Total = 23  
Accuracy =0.65  
Precision = 0.65  
Recall = 1  
Specificity = 0.0  
F-measure = 0.79  
NVP = nan  
FPV = 0.35

Classification report

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Figure 4 Sigmoid Classification report

Comparing the actual results to the results computed by classification report, they are the same results.

**b) ROC curves**

**The *receiver operating characteristic curve* is a method for evaluating analysis strategies. It represents the dependence of efficiency with error rate for different parameter values. A curve close to the diagonal indicates a random process: Values close to the diagonal mean an equal hit rate and false positive rate, which corresponds to the expected hit frequency of a random process. Particularly the sigmoid curve in the first plot is below the diagonal. Sigmoid performs better with a high data set.** After tuning the hyperparameter all curves, included the sigmoid curve, are above the diagonal. We now have more accurate results.

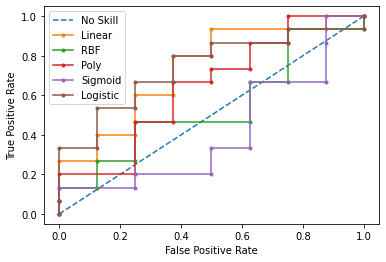


Figure 5 ROC-curve before hyperparameter tuning

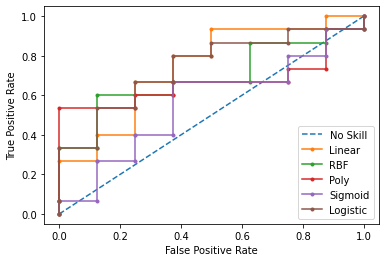


Figure 6 ROC-curve after hyperparameter tuning

**b) Quality results**

**Comparing the accuracies, the linear model with 70% accuracy, turns out to be the best.**

**c) Qualitative Results**

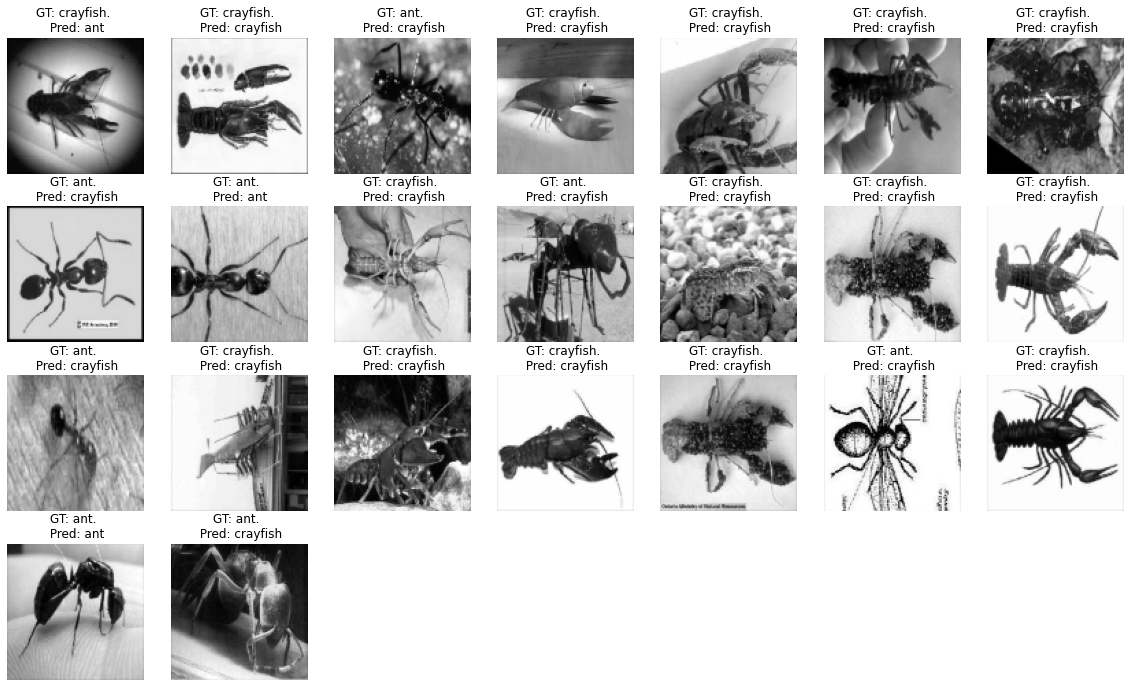


Figure 7 Sample of test images of linear model

Plotting a few sample results of the linear model we see 18 of 23 true predictions.