ML and PR Project Report – Part 7

The theme of the current laboratory is Support Vector Machines, or SVMs. We are mainly going to focus on the simple linear variant, and the kernel ones, with polynomial of degree 2 kernel, and Radial Basis Function, or RBF, kernel.

# Simple (linear) SVM

The first analysis we are conducting is a simple linear SVM classifier. The regularization term for the bias is kept constant, at a value of 1, throughout the analysis, and the application considered has an effective prior of 0.1. What we are going to focus on is the performance, in terms of minimum and actual DCF, of the classifier, on the aforementioned application, when varying the cost value C, which can also be seen as a regularization coefficient, for which greater values imply weaker regularization.

As the threshold used when actually classifying a certain sample with a label is arbitrarily chosen to be 0, we shouldn’t expect our results to exhibit signs of good calibration, and that is actually the case.

Immagine che contiene testo, Diagramma, schermata, linea

Descrizione generata automaticamente

By looking at the curves above we may see that, while having too strong of a regularization may lead to completely unsatisfactory results for both actual and minimum DCF, decreasing it even slightly brings the minimum DCF results to their actual best, meaning that the performance remains constant even through such weak regularization that C becomes 1. As anticipated, the performance of the actual DCF doesn’t start improving until lower amounts of regularization, to then reach its best results with much worse values that the minimum DCF counterpart.

With linear SVM being, as the name suggests, a classification model that is only able to generate linear decision boundaries, we may expect its results to be on par with the other linear models we have analyzed so far, and that actually is the case, as the values for the minimum DCF sit slightly above the 0.35 mark.

The type of objective of the three linear models seen so far (Tied, Logistic Regression, and this one) are quite different: Tied takes decisions based on the distribution of the samples, Logistic Regression on that of the classes, while SVM does not rely on a probabilistic interpretation at all. With that in mind, we could wonder if such differences may lead to different results in performance, but this doesn’t seem to be the case for the minimum DCF, while it’s clear that the actual DCF performance is subpar for the SVM case.

One more thing that we may try is preprocessing. In this context we decided to simply apply pre-centering, while keeping in mind that the structure of the objective does not allow for invariance of the decisions and results with respect to linear transformations of the data. That being said, as we have already observed in the previous parts of the project, the data does not seem to deviate much from its center already, meaning that we should not expect great performance differences compared to the non-centered version of the classifier.

Immagine che contiene testo, linea, Diagramma, schermata

Descrizione generata automaticamente

The results are basically identical, as expected.

# Polynomial (degree 2) Kernel SVM

Just like with Logistic Regression, the dataset can be transformed to accommodate for non linear decisions, at the cost of an increase in number of computations. Thanks to the possibility of formulating the SVM problem also in its dual form though, we are able to perform the same type of classification that we do in the primal form, using just dot products between data points, and thus being able to use kernel functions.

The first type of kernel we try to apply is a degree 2 polynomial, for which we set the bias term to be zero, since the polynomial kernel already accommodates for its own bias term c, which we set to 2.

Based on what we have learned about the way the data is distributed in the previous labs, it’s safe to expect that a non-linear decision is going to be better suited for our dataset.

Immagine che contiene testo, linea, Diagramma, schermata

Descrizione generata automaticamente

The results are better than the linear version of the SVM. The behavior of the minimum and actual DCF are the same as the linear SVM with regards to their relation with normalization (except for a slightly more observable increase in the cost of the minimum DCF when the normalization is too low), but both reach lower values by a whole 0.1 DCF points, and the actual DCF also has a much faster cost descent.

Although the actual DCF’s best value has lowered by a whole 0.1 in cost, it is still subpar compared to the actual DCFs seen for the Tied and Logistic Regression models, for which the actual DCF’s best values did perform worse that the minimum one, but not by this big of a margin.

The best results of the minimum DCF are the same we had for the Quadratic Logistic Regression model.

# RBF Kernel SVM

Another kernel we want to try is a Gaussian Radial Basis Function (), for which we set bias to 1, and try out different . The latter parameter allows us to shape the decision boundaries by punishing more or less some distance between the point that needs to be classified and the Support Vectors, or better, by defining how far the influence of a single Support Vector reaches, with low values meaning ‘far’ and high values meaning ‘close’.

Again we also vary the normalization coefficient.

Immagine che contiene testo, schermata, linea, diagramma

Descrizione generata automaticamente

The results of the minimum DCF are amazing, with the and versions being the first models to beat the 0.2 cost so far, and the others reaching slightly better results than the former rank 1 model (Quadratic Logistic Regression). Since the best result was obtained with the greater values of , we may guess that our data benefits from harsher decisions, i.e. decisions that lower the score for a sample much quicker in function of the distance of that point to the Support Vector.

The relation between the actual DCFs and the normalization factor doesn’t change much with respect to the previous models, except the differences in descent “speed” and minimum hit. For the minimum DCF instead, we observe that the results start at acceptable values even for excessive normalization, and they mostly get better when lowering it. To be precise, there are some spikes in cost when normalization is too low, especially for the strongest application.

For some models (e.g. the one), there are some values of C that lead to measuring good calibration (see the C = 1 points), but in general, this keeps being subpar as for the other SVM models.

# Conclusion

Linear SVM performs like the other linear models, in term of minimum DCF, but does worse when looking at the actual DCF, suggesting a calibration problem. Linear SVM does not benefit from pre-centering.

Nonlinear decision boundaries yield much better results, with the RBF one being the best model seen so far, especially with the more rigid (i.e. greater) gammas, and the Quadratic following the Quadratic Linear Regression performance. Unfortunately, in terms of calibration, they still perform badly.