

Assignment 1

by: Max Verhagen (a1725532)

Support-vector machines also known as 'SVMs' is a supervised learning model used to categorise its labelled training data also known as classes, in this assignment the two classes we are given are zero and one, hence known as a binary class. The Support-vector machines functional classification is done by attempting to find the optimal hyperplane that separates data points by class.

For any given hyperplane it can be expressed as $f(x) = w * x + b = 0$. In addition to the creation of the hyperplane a margin is also formed by $f(X) = -1$ and the other side being $f(x) = 1$. This margin is defined by the support vectors and the distance they are from the separating hyperplanes. Therefore in short support vectors are points of data within a class that is closest to the hyperplane and define the margin.

There are two different types of margins; these being hard margin and soft margin. Traditionally speaking hard margin is used in an attempt to find completely linearly separable data without the tolerance for errors. The fault with hard margin support-vector machines is that single outliers can significantly affect the margin leading to drastic overfitting.

In contrast to hard margin, soft margin allows for misclassification of data points in an attempt to minimize overfitting and maximize the margin. Regularization weight commonly also defined as C in regard to Support-vector machines soft margins is. The larger the Regularization weight the less tolerance there is for outliers, receiving a larger penalty when something is misclassified. In essence setting a very large regularization weight is the same as using hard margin.

Larger margin size allows Support-vector machines to create decisions more confidently. Therefore finding a solution with the largest margin is optimal as it provides larger certainty with decision making and gives classification a safety margin for creating errors with misclassification.

Within Linear Programming through the concept of duality, it is stated that within every linear problem the original being known as the 'Primal' there is another problem related to it known as the 'Dual' and therefore can be derived from it.

This process starts with the original linear problem being formulated with all variables being non-negative, this creates what is known as the standard form. In essence primal is solving the problem of maximization $c^t x$ where $Ax = b, x \geq 0$ and dual is minimization of $b^t y$ where $A^t y \geq c$ with y being unrestricted.

Duality gap is the difference between Primal least object value and the greatest objective value of dual. In the case that the difference between the two is non-zero this is known as weak duality. Weak duality shows that we haven't been able to find an optimal solution or in any case prove that the solution is optimal. On the other hand, when both have a finite optima that has no gap, this is known as strong duality and confirms that the optimal solution has been found.

Experiments

Primal

Bias : 0.00011326512142805203
Max weight: 4.1111178874500376e-05
Min weight: -4.1565539256661785e-05
Mean weight: 6.8940849617113374e-09

Dual

Bias : 2.1282549135643567
Max weight: 0.8702413423207205
Min weight: -0.6703565210848794
Mean weight: -0.0015376347295953147

When comparing the two results we find the bias and weight for the Primal results are significantly smaller than the Dual results. This is due to the calculation of optima via the use of minima and maxima.

Coded dual

Bias: 2.1282549135643567

Max weight : 0.8702413423207205
Min weight: -0.6703565210848794
Mean weight : -0.0015376347295953147

Max alpha : 0.9999999988336159
Min alpha : 1.1185854912389075e-06
Mean alpha : 0.8210018732207427

Libsvm dual

Bias: 1.01287616

Maximum weight: 0.4157055965796455
Minimum weight: -0.31945708683942264
Mean weight : -0.00017700718023295177

Errors percentage

Calculated via the formula: Wrong answer / Total answers * 100

Coded prime

Train: 13.95%

Test: 7.27%

SVC prime

Train: 2.38%

Test: 2.53%

Coded dual

Train: 2.21%

Test: 3.14%

SVC dual

Train: 2.38%

Test: 2.53%

The reason that we have a margin of error when testing the support vector machines with their own training data is due to the use of soft margins. If there was an increase to the Regularization weight or even change to hard margin it would be likely observed that the Errors percentage for testing the training data would decrease likely to zero. However, the errors percentage for the test data would likely drastically increase due to overfitting.

The following are confusion matrix this provides insight into how the test is being categorised and provide insight into where the support vector machine maybe going wrong in terms of categorization. This comparison was only done for the test data.

Prime	Predicted : 0	Predicted: 1
Actual : 0	681	78
Actual : 1	31	710

Prime svc	Predicted : 0	Predicted: 1
Actual : 0	736	23
Actual : 1	15	726

Dual	Predicted : 0	Predicted: 1
Actual : 0	733	26
Actual : 1	21	720

Dual svc	Predicted : 0	Predicted: 1
Actual : 0	736	23
Actual : 1	15	726