ASSESSMENT COVER SHEET

Assignment Details

Course: 3920_COMP_SCI_X_0009

Semester/Academic Year: Semester 2 - 2019

Assignment title: Assignment 1

Assessment Criteria

Assessment Criteria are included in the Assignment Descriptions that are published on each course's web site.

Plagiarism and Collusion

Plagiarism: using another person's ideas, designs, words or works without appropriate acknowledgement.

Collusion: another person assisting in the production of an assessment submission without the express requirement, or consent or knowledge of the assessor.

Consequences of Plagiarism and Collusion

The penalties associated with plagiarism and collusion are designed to impose sanctions on offenders that reflect the seriousness of the University's commitment to academic integrity. Penalties may include: the requirement to revise and resubmit assessment work, receiving a result of zero for the assessment work, failing the course, expulsion and/or receiving a financial penalty.

DECLARATION

I declare that all material in this assessment is my own work except where there is clear acknowledgement and reference to the work of others. I have read the University Policy Statement on Plagiarism, Collusion and Related Forms of Cheating:

http://www.adelaide.edu.au/policies/?230

I give permission for my assessment work to be reproduced and submitted to academic staff for the purposes of assessment and to be copied, submitted and retained in a form suitable for electronic checking of plagiarism.

Max Verhagen   4/09/2019

SIGNATURE AND DATE

Assignment 1

by: Max Verhagen (a1725532)

Support-vector machines also known as 'SVMs' is a supervised learning model used categorise its labelled training data also known as classes, in this assignment the two classes we are given are zero and one, hence know 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. Therefor 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 attempted 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 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 outliers, receiving a larger penalty when something is misclassified. In essence a 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. Therefor finding a solution with the largest margin is optimal as provides larger certainty with decision making and gives classification a a safety margin for creating errors with misclassification.

Within Linear Programming thought the concept of duality, it is stated that within every linear problem the original being know as the 'Primal' there is another problem related to it know was the 'Dual' and therefor 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 defence 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 know 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%

Coded dual

Train: 2.21%

Test: 3.14%

SVC prime

Train: 2.38%

Test: 2.53%

SVC dual

Train: 2.38%

Test: 2.53%

The reason that we a margin of error when testing the support vector machines with there 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 decrees 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 compression 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