Homework 4

Evan Sidrow

November 10, 2017

1 Support Vector Machines

1.1 Programing questions

See svm.py

1.2 Analysis

1.2.1 Use the Sklearn implementation of support vector machines to train a classifier to distinguish 4's from 9's

Done.

1.2.2 Experiment with linear, polynomial, and RBF kernels. In each case, perform a Gridsearch to help determine optimal hyperparameters for the given model. Comment on the experiments you ran and the hyperparameters you found.

Firstly, I had to restrict the number of training instances to 1000 due to run-time constraints. In addition, I used parameters that were in powers of ten for the grid search (i.e C $\in \{0.01, 0.1, 1, 10, 100\}$). I also ran the grid search again with more parameters if any of the optimal parameters returned were at the edge of my search space. This ensured that I wasn't looking in the wrong place for good parameter values. After running GridSearchCV on all three kernels with 1000 instances in the training set, I got the following optimal values:

Table 1: Optimal hyperparameter values for different kernels in a SVM

Kernel	Linear	RBF	Polynomial
C	0.1	10	1000
γ	-	0.01	-
p (degree)	-	-	2

1.2.3 Comment on classification performance for each model for optimal parameters by testing on a hold-out set or performing cross-validation.

I used a hold-out set to test my classification performance. For each SVM with the optimal hyperparameter values, I got the following results:

Table 2: Test and Training accuracy for different kerneled SVMs with optimal hyperparameters

Kernel	Linear	RBF	Polynomial
Training Accuracy	0.987	1.0	1.0
Test Accuracy	0.959	0.976	0.972

1.2.4 Give examples (in picture form) of support vectors from each class when using a polynomial kernel.

The support vectors for this set are the 4's and 9's that look most like each other. Note that there were a total of 231 support vectors after training the polynomial SVM on 1000 training instances. Here are a couple of those support vectors:

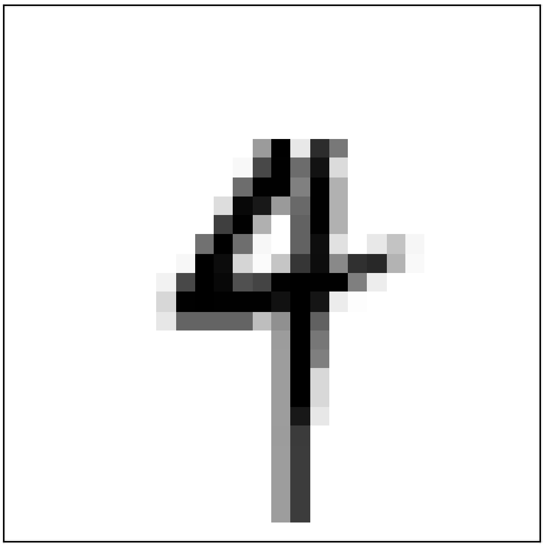

Figure 1: A picture that is a support vector labeled as a 4

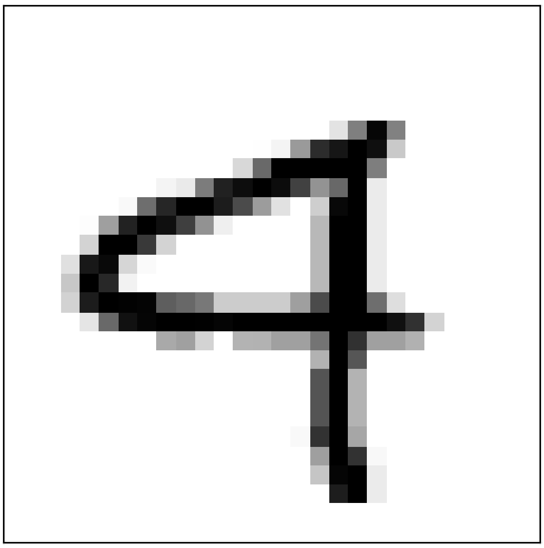

Figure 2: A picture that is a support vector labeled as a 9

Yep, that four looks like a nine (the top of it is almost closed). The 4 in figure 2 was classified as a nine, since its top is closed and round-ish. I could be fooled into thinking that it is a nine.

2 Learnability

For the concept class C of all triangles where each vertex has the form $(i, j) \in [0, 99]X[0, 99]$, we can have our hypothesis class H be the same set. In that case, then we will have to pick 3 vertices from a 100-by-100 grid, and there are $\binom{100^2}{3}$ ways to do that, so $|H| = \binom{100^2}{3}$.

Now, since this a finite, consistent hypothesis class, then we have a formula for the minimum number of training examples needed to achieve a generalization error of less then ϵ with probability greater than $1 - \delta$ (here $\epsilon = 0.15$ and $\delta = 0.05$).:

$$m \geq \frac{1}{\epsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

$$m \geq \frac{1}{0.15}\left(\ln\binom{100^2}{3} + \ln\frac{1}{0.05}\right)$$

$$m \geq \frac{1}{0.15}\left(\ln\binom{100^2}{3} + \ln\frac{1}{0.05}\right)$$

$$m \geq 192.2313$$

So we need at least 193 training examples to ensure that there is 95% chance that our classifier has error no larger than 0.15.

3 VC Dimension

3.1 Circle centered at the origin

For the VC Dimension of a circle centered at the origin, **The VC dimension is 2**.

We first prove that the VC dimension is at least 2. For this, consider two points in polar coordinates, $x_1 = (r_1, \theta_1)$ and $x_2 = (r_2, \theta_2)$, where $r_1 > r_2$. Now, consider these two cases:

Case 1: **x_1 and x_2 have different labels**. In this case, we prove that $r = \frac{r_1+r_2}{2}$ separates these points. First, it is clear that $r_1 > \frac{r_1+r_2}{2} > r_2$, so r_1 lies outside the circle, and r_2 lies inside the circle. Then, we say everything inside the circle is labeled the same as x_2 and everything outside the circle is labeled the same as x_1.

Case 2: **x_1 and x_2 have the same label.** In this case, we prove that $r = r_1 + 1$ separates these points. First, it is clear that $r_1 + 1 > r_1 > r_2$, so both r_1 and r_2 lie inside the circle. Then, we say everything inside the circle is labeled the same as x_2 and x_1 and everything outside the circle is the other label.

Now, to prove that the VC dimension is no more than three, we show that there is no set of points x_1, x_2, and x_3 that can be shattered by a circle. We will remain in polar coordinates.

Without loss of generality, let $r_1 \geq r_2 \geq r_3$. Then, label x_1 and x_3 as positive and x_2 as negative. Now suppose we have a circle described by $r = r_0$.

- If $r_0 \geq r_2$ and we classify everything on the boundary or inside the circle as positive, then we misclassify x_2.
- If $r_0 < r_2$ and we classify everything on the boundary or inside the circle as positive, then we misclassify x_3.
- If $r_0 > r_2$ and we classify everything on the boundary or outside the circle as positive, then we misclassify x_1.
- If $r_0 \leq r_2$ and we classify everything on the boundary or outside the circle as positive, then we misclassify x_2.

So we have shown that the VC dimension is at least 2, but less than 3, so the VC dimension MUST be 2.

3.2 EXTRA CREDIT: Circle centered anywhere

For this question, We prove that the VC dimension is at least 3, but less then 4, so **the VC dimension of a circle centered anywhere MUST be 3.**

First, we prove that the VC dimension is at least 3. Consider any set of th3 points that make up a triangle in 2D space. We note that any three points are linearly separable by a line, and the limiting case of a circle as the radius goes to infinity is a line. Therefore, not matter what the labels of these three points, a line (and therefore a circle with a sufficiently large radius) can correctly classify them.

Next, we prove that the VC dimension can be no larger than 4. Here, we consider two cases:

Case 1: **One point is inside (or on the boundary of) the (possibly degenerate) triangle described by the other three points.** If this is the case, label the point on the convex hull of the others as negative and the other three points as positive. We note that both circles and triangles (even degenerate ones) are convex, so any circle that contains the three positive points must contain the (possibly degenerate) triangle that they are a part of. However, that means that the circle must also contain the negatively classified point inside the triangle, so this point is misclassified.

Case 2: **All four points create a convex quadrilateral.** This case is more difficult. Label each point A,B,C, and D, where A and C are opposite one another and B and D are also opposite one another. We know that the sum of the angles of this quadrilateral is 360 degrees, so there is one pair of angles opposite to one another whose sum is ≤ 180 degrees. Without loss of generality, assume these points are A and C label them as positive and B and D as negative.

By way of contradiction, assume that there is a circle that separates these points. Then, we can draw a circle that is contained inside this circle which has A and C on its boundary. Now, we have that two angles (A and C) on the edge of a circle sum to ≤ 180 degrees.

However, the sum of opposite angles of a quadrilateral inscribed in a circle must add to 180 degrees. If this sum is less than 180, then the other points must be inside the circle. This means that B and D are either on the edge of the circle (if angles A and C add to 180 degrees), or either B or D is inside the circle (if angles A and C add to less than 180 degrees). This means that B or D is misclassified, so we have a contradiction!

Here is a picture that might make sense:

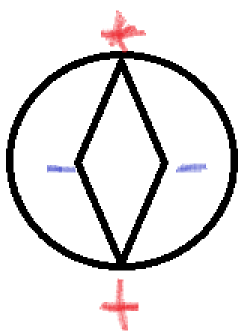

Figure 3: The positive points have angles that sum to less the 180 degrees, so at least one negative point must be in the circle

Therefore, the VC dimension of this circle is at least 3, but less than 4, so the VC dimension must be 3.