

# B555: Assignment 4

Pawan Patel

December 9, 2015

1. a)  $f(x, y) = x^2 - y^2$  on  $x^2 + y^2 = 4$  yields the system  $\{2x = \lambda 2y, -2y = \lambda 2y\} \implies \lambda = 1, -1$ . If  $\lambda = 1 \implies y = 0 \implies x = 2, -2$ . If  $\lambda = -1 \implies x = 0 \implies y = 2, -2$ . This produces four critical points of  $(2, 0), (-2, 0), (0, 2), (0, -2)$ , with a maximum of 4 at the first two points and a minimum of -4 at the second two points.

b)  $f(x, y) = x^2 y - \log x$  on  $x + 2y = 0$ . This function has no maximums or minimums. Indeed, we have that  $y = -\frac{1}{2}x$  and  $x > 0$  and so we may write  $f$  as a function of one variable as  $-\frac{1}{2}x^3 - \log(-\frac{1}{2}x)$  whose domain is  $(0, \infty)$ , on which it is continuously differentiable. We can then see that as  $x \rightarrow 0$  this function goes to  $-\infty$  and as  $x \rightarrow \infty$  this function goes to  $\infty$ . Thus, there are no maximums or minimums.

c)  $f(x, y) = x^2 + 2xy + y^2 - 2x$  on  $x^2 - y^2 = -1$ . This function has no maximums or minimums. We can set  $x^2 = y^2 - 1$  and compute  $f$  on two branches. For positive  $y$ ,  $f(y) = 2y^2 + (2y - 2)\sqrt{y^2 - 1} - 1$ , which goes to  $\infty$  as  $y \rightarrow \infty$ . For negative  $y$ ,  $f(y) = 2y^2 - (2y - 2)\sqrt{y^2 - 1} - 1$ , which goes to  $-\infty$  as  $y \rightarrow \infty$ .

2. (a)-(c): For this problem, I implemented three types of kernels based on a fixed number of centers. The first is the kernel proposed in part (a),  $k(x, x_i) = \langle x, x_i \rangle$ . The second is a polynomial kernel where  $k(x, x_i) = (1 + \langle x, x_i \rangle)^d$  and the third is a radial kernel where  $k(x, x_i) = \exp(-\|x - x_i\|_2^2)$ .

The implementations requires various choices to be made. The first is the choice of centers. For this problem, I chose to use 10 centers for convenience, but one can easily change the number of centers for analysis. Ideally, such a choice would be made with some apriori information on the data. Next, to choose these centers, I use the k-means clustering algorithm from sklearn. I primarily chose this because it a clustering algorithm I am familiar with and because it is readily available. Choice of centers is an important consideration. In addition, for all three, I added a regularizing term for the linear regression step to avoid invertibility issues. Finally, for the polynomial kernel I chose a degree of 4 and for radial kernel I chose a  $\sigma$  of 2. Each of these can

easily be changed in the implementation, though there does not seem to be significant differences in doing so.

All of the approaches do reasonably well, but only the normal kernel regression beats linear regression. Polynomial and Radial Kernels come close to vanilla linear regression, but do not beat it. For the polynomial kernel, a lower degrees tend to do better, which seems consistent with the fact that the normal kernel works the best. For radial kernels, a  $\sigma$  of 2 or 3 seems to work best, but the difference doesn't seem to be statistically significant.

3. For this problem, I chose to compare linear regression and logistic regression on the susy dataset. I use 10 splits of the 100,000 data points for training and testing sets. For each training split, I further use 10 folds, along which I train using both Linear Regression and Logistic Regression using all parameters, selecting the parameters for each that yield the highest average accuracy across the 10 folds. For Linear Regression, I adjust the step size for gradient descent from .01 to 5. For Logistic Regression, I adjust two parameters, the step size from .01 to 5 and the number of iterations from 5 to 500. This yields 10 data points for the average accuracy for both types of regression. I get the following results: Linear Regression: [ 59.6 60.35 59.5 60.15 60.46 60.2 60.68 59.89 60.08 60.31] and Logistic Regression: [ 54.2 53.61 54.57 53.55 53.99 53.02 54.09 54.26 54.14 54.82].

Based on these numbers, I hypothesis that Linear Regression out performs Logistic Regression. Therefore, I do a Welch's two sample t-test with the null hypothesis that the mean accuracy from Linear Regression is lower than the mean accuracy for Logistic Regression. This yields a p value of  $5.3 \times 10^{-16}$ . Therefore, I reject the null hypothesis and conclude that the mean accuracy from Linear Regression is higher than the mean accuracy for Linear Regression. In particular, Linear Regression is the better performing algorithm. (see test.r in folder for code for the test).