Technical Report 4: Regularization

1 Normal Equations for Linear Regression

The 5-order linear regression results with different regularization parameters are illustrated in Fig. 1. It can be observed that, the boundary well separates the positive data samples from the negative ones. As shown in Fig. 1 (a), when $\lambda = 0$, the regularization is disabled, such that the resulting curve perfectly fits the given data points. In this overfitting case, the regression results actually cannot show a general trend and thus may not be able to make an accurate prediction. By tuning λ, we have more control over the data fitting. In Fig. 1 (b), we have a "simplified" fitting curve under $\lambda = 1$. The curve does not fit the training data perfectly, but still demonstrate the trend of the data. In Fig. 1 (c) where $\lambda = 10$, the curve is too "smooth" to fit the training data. This case is so-called "underfitting".

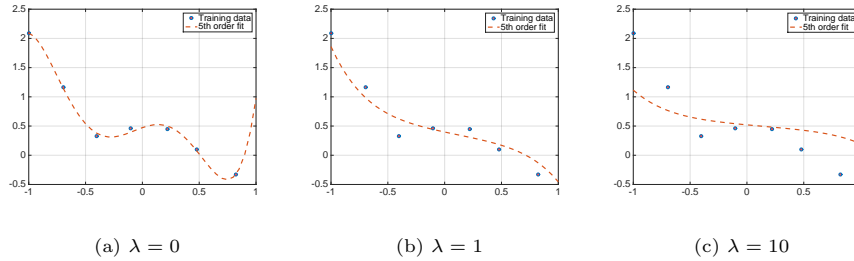| (a) $\lambda = 0$ | (b) $\lambda = 1$ | (c) $\lambda = 10$ |

Figure 1: Regularized linear regression under different regularization parameters.

We also give each element of θ under different settings of λ in Table 1 Lager λ exerts stronger penalties to the parameters (of the hypothesis function), especially the larger ones. Therefore, we have the norm of θ decreased with increasing λ.

2 Newton's Method for Regularized Logistic Regression

We apply Newton's method to non-linear regularized logistic regression with different λ's. The results are visualized in Fig. 2. When $\lambda = 0$ (such that we have a non-regularized logistic regression), the positive and negative training data are separated accurately by the resulting decision boundary. There is even

Table 1: θ under different regularization parameters

	$\lambda = 0$	$\lambda = 1$	$\lambda = 10$
θ_0	0.4725	0.3976	0.5205
θ_1	0.6814	-0.4207	-0.1825
θ_2	-1.3801	0.1296	0.0606
θ_3	-5.9780	-0.3975	-0.1482
θ_4	2.4418	0.1753	0.0743
θ_5	4.7374	-0.3394	-0.1280
$\|\theta\|$	8.1690	0.8098	0.5931

an island negative region inside the positive region. Similar to the above case of regularized linear regression, the decision boundary is too precise to reflect the general trend of the training data. In another word, the decision boundary over-fits the training data in this case. The $\lambda = 1$ plot shows an simpler decision boundary which still separates the positives and negatives fairly well. On the other hand, $\lambda = 10$ is probably too high of a regularization parameter, as the boundary decision does not follow the data so well on in the lower-left region.
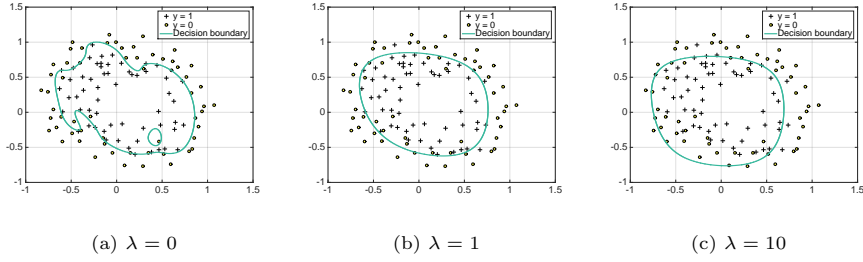

(a) $\lambda = 0$ (b) $\lambda = 1$ (c) $\lambda = 10$

Figure 2: Regularized logistic regression under different regularization parameters.

We also report the norms of θ's under different λ's in Table 2. The norms of θ decrease as λ increases. There are also visible fitting changes in the corresponding plots.

Table 2: θ under different regularization parameters

	$\lambda = 0$	$\lambda = 1$	$\lambda = 10$
$\|\theta\|$	7.173×10^3	4.240	0.9384