Regularization example

We'll comence here by expanding a bit on the relation between the "effective" number of parameter choices and regularization discussed in the lectures. We do this in the context of a simple 1-dim logistic regression model

$$P(y = 1|x, \mathbf{w}) = g\,(\,w_0 + w_1 x\,) \tag{1}$$

where $g(z) = (1 + \exp\{-z\})^{-1}$. We will assume here that $x \in [-1, 1]$.

To understand regularization in this context, we'll try to carve up our parameter space $\mathbf{w} = [w_0, w_1]^T \in \mathcal{R}^2$ into regions such that our loss (or log-probability) is roughly constant within each region. This will help us determine how many "effective" parameter choices we really have. Ideally, the regularization that we impose would directly limit (our estimate of) the number of parameter choices.

A bit more precisely, we want the log-probability $\log P(y|x, \mathbf{w})$ to vary by no more than ϵ within each region in the parameter space. To find such regions, we first examine how the log of the logistic function varies as a function of its input: (this result will be useful to you later on)

$$\frac{\partial}{\partial z} \log g(z) = \frac{1}{g(z)} \frac{\partial}{\partial z} g(z) = \frac{1}{g(z)}\, g(z)(1 - g(z)) = 1 - g(z) \tag{2}$$

In other words, since $g(z) \in [0, 1]$ (probability), the derivative here is also bounded by 1. The function $\log g(z)$ therefore varies at most linearly with slope 1 as a function of z, or

$$|\log g(z) - \log g(z')| \leq |z - z'| \tag{3}$$

for any two points z and z'. To use this result, we define $z = w_0 + w_1 x$ and $z' = w_0' + w_1' x$ for any $x \in [-1, 1]$. This gives

$$\begin{aligned} |\log P(y = 1|x, \mathbf{w}) - \log P(y = 1|x, \mathbf{w}')| &= |\log g(z) - \log g(z')| \leq |z - z'| && (4) \\ &= |(w_0 - w_0') + (w_1 - w_1')x| && (5) \\ &\leq |(w_0 - w_0')| + |(w_1 - w_1')x| && (6) \\ &\leq |(w_0 - w_0')| + |(w_1 - w_1')| && (7) \end{aligned}$$

since $|x| \leq 1$ by assumption. So, whenever $|w_0 - w_0'| + |w_1 - w_1'| \leq \epsilon$, the corresponding losses or (negative) log-probabilities are also bounded by ϵ. We can therefore carve up the parameter space by finding discrete points $\mathbf{w}^{(i)}$ and regions around them such that

$$|w_0^{(i)} - w_0'| + |w_1^{(i)} - w_1'| \leq \epsilon \tag{8}$$

whenever $\mathbf{w}'$ belongs to the i^{th} region. These regions are shown in Figure 1 for $\epsilon = 0.4$. We have also included in the Figure the area limited by the Euclidean norm of the parameter vector, $\|\mathbf{w}\|^2$. Increasing the the limit $\|\mathbf{w}\|^2$ clearly incorporates more "choices" and it

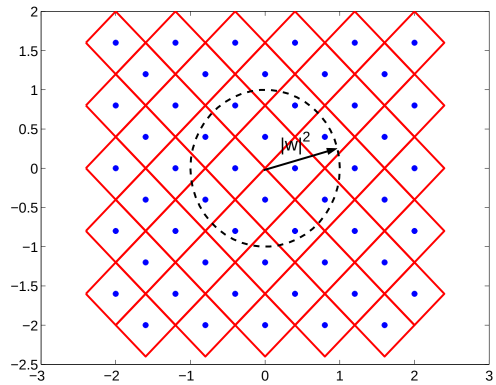

Figure 1: Regions in the parameter space corresponding to roughly constant losses for simple logistic regression model.

makes sense to use this type of norm in regularization, i.e., in limiting the effective number of parameter choices.

Note: Explicitly finding the regions as we have done here is merely a conceptual device in understanding (and analyzing) estimation methods. We never have to identify such regions nor the discrete “effective” parameter choices in practice.