Project 4 - Derivations

Taylor B. Morris

December 15, 2016

For kernel logistic regression, we start with the formulas for regularized linear logistic regression,

$$P(y|x) = \frac{1}{1+\exp(-yi(w^T x + b))}$$

which gives

$$w^* = \min_w \frac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{n} \ln(1+\exp(-y(w^T x_i + b)))$$

Let $g(\xi) = \ln(1+e^{\xi})$. Then we have

$$\min \frac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{n} g(\xi_i)$$

With the constraint

$$\xi_i = -y_i(w^T x_i + b)$$

We can now form the Lagrangian as

$$L = \frac{\lambda}{2}\|w\|^2 + \sum_i g(\xi) + \sum_i \alpha_i \left[-\xi - y_i(w^T x_i + b)\right]$$

Taking the partial derivative with respect to w, we get

$$\Delta_w L = \lambda w - \sum_i \alpha_i y_i x_i$$

Next, setting to 0 and solving for w,

$$w = \frac{1}{\lambda}\sum_i \alpha_i y_i x_i$$

Next, by taking the derivative with respect to b, we obtain

$$\Delta_b L = \sum_i \alpha_i y_i = 0$$

Now, taking the partial derivative with respect to ξ, we get

$$\Delta_\xi L = g'(\xi_i) - \alpha_i$$

Setting this to 0,

$$g'(\xi_i) = \alpha_i$$

Remember that, because

$$g(r) = \ln(1 + e^r)$$

We have

$$g'(r) = \frac{e^r}{1 + e^r}$$

Now, multiplying up, we have

$$g'(r) + g'(r)e^r = e^r$$

$$e^r(1 - g'(r)) = g'(r)$$

$$e^r = \frac{g'(r)}{1 - g'(r)}$$

$$r = \ln \frac{g'(r)}{1 - g'(r)} = (g')^{-1}(g'(r))$$

Now, we plug our α and ξ back in, and we get

$$\xi = g'(\alpha) = \ln \frac{\alpha}{1 - \alpha}$$

Making a new function from the terms in the Lagrangian that feature ξ, we have

$$G(\alpha_i) = g(\xi_i) - \alpha_i \xi_i$$

To make $G(\alpha_i)$ completely in terms of α, we can

$$G(\alpha_i) = \ln(1 + \frac{\alpha_i}{1 - \alpha_i}) - \alpha_i \ln \frac{\alpha_i}{1 - \alpha_i}$$

Now, we can plug some of our information back into the Lagrangian to get

$$L = \frac{\lambda}{2}(\frac{1}{\lambda}\sum_i \alpha_i y_i x_i)^2 + \sum_i G(\alpha_i) - \sum_i \alpha_i y_i (\frac{1}{\lambda}\sum_j a_j y_j x_j x_i + b)$$

Now, because we have the new constraint $\sum_i \alpha_i y_i = 0$, we can finish our form as

$$L = \frac{1}{2\lambda}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i G(\alpha_i) - \frac{1}{\lambda}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Now, with our kernel $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, we have

$$\min_\alpha \frac{1}{2\lambda}(\alpha * y)^T K(y * \alpha) + \sum_i G(\alpha_i)$$

Putting our w back into our probability, we get

$$p(y_c|x) = \frac{1}{1 + \exp(-y_c((\sum_i a_i y_i x_i)x + b)}$$

Using our test kernel here, we can see

$$p(y_c|x) = \frac{1}{1 + \exp(-y_c((a * y)^T K_t + b)}$$