Kernelized Logistic Regression

$$x \in \mathbb{R}^d \longrightarrow \phi(x) \in \mathbb{R}^D \qquad D \gg d$$

Can we work with inner products $\phi(x_1)\cdot\phi(x_2)$ only ?
Since D is large (often larger than number of examples n), we will consider regularized log loss.

Primal: $$\min_{w} \sum_{i=1}^{n} \log\left(1+e^{-\langle w, x_i\rangle y_i}\right) + \frac{\lambda}{2}\|w\|^2$$

Equivalently, lets write this as constrained optimization and use Lagrangian duality (you can also try to use Fenchel aka function duality if you prefer).

$$\min_{w,\{z_i\}} \sum_{i=1}^{n} \log\left(1+e^{z_i}\right) + \frac{\lambda}{2}\|w\|^2$$

$$\text{s.t.} \quad z_i = -\langle w, x_i\rangle y_i$$

Lagrangian $$\mathcal{L} = \sum_{i=1}^{n} \log\left(1+e^{z_i}\right) + \frac{\lambda}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i\left(z_i + \langle w, x_i\rangle y_i\right)$$

↳ Lagrange multipliers (don't need to constrain sign since corresponding constraint is equality)

$$\frac{\partial \mathcal{L}}{\partial z_i} = \frac{e^{z_i}}{1+e^{z_i}} - \alpha_i = 0 \Rightarrow z_i = \log\frac{\alpha_i}{1-\alpha_i}$$

$$\frac{\partial \mathcal{L}}{\partial w} = \lambda w - \sum_{i=1}^{n}\alpha_i x_i y_i = 0 \Rightarrow \underline{w = \frac{1}{\lambda}\sum_{i=1}^{n}\alpha_i x_i y_i}$$

This is the key observation which helps us connect primal variables w with dual variables α

We can now proceed to derive the dual (see below*) or simply plug-in this form into the logistic model as shown in lecture:

$$P(Y=y\mid X) = \frac{1}{1+e^{-\langle w, x\rangle y}} = \frac{1}{1+e^{-\frac{1}{\lambda}\sum_{i=1}^{n}\alpha_i\langle x_i, x\rangle y_i y}}$$

$$= \frac{1}{1+e^{-\sum_i \alpha_i'\langle x_i, x\rangle y}}$$

α_i' absorbs λ & y_i

For $\langle\phi(x_i),\phi(x)\rangle$, replace with $K(x_i, x)$

We can now just optimize our loss by running gradient descent on α.

(Note $\|w\|^2$ can also be written using kernels & α)

* If we want to get dual problem via maxmin formulation, lets replace z_i & w using dual variables in the Lagrangian

$$\min_{w,\{z_i\}} \mathcal{L} = -\sum_{i=1}^{n}\log(1-\alpha_i) + \frac{1}{2\lambda}\sum_{ij}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle - \sum_{i=1}^{n}\alpha_i\log\frac{\alpha_i}{1-\alpha_i}$$
$$- \frac{1}{\lambda}\sum_{ij}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle$$

Dual problem $$\max_{\alpha}\min_{w,\{z_i\}} \mathcal{L}$$

$$= \max_{\alpha} \sum_{i=1}^{n} H(\alpha_i) - \frac{1}{2\lambda}\sum_{ij}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle$$

↳ entropy

$$H(\alpha_i) = -\alpha_i\log\alpha_i - (1-\alpha_i)\log(1-\alpha_i)$$

If $x \to \phi(x)$, we can use kernels in dual problem

$$\max_{\alpha} \sum_{i=1}^{n} H(\alpha_i) - \frac{1}{2\lambda}\sum_{ij}\alpha_i\alpha_j y_i y_j K(x_i, x_j)$$