# Pset 3

Noah Toyonaga

December 8, 2022

## 1 Problem 1

### 1.1 Explicit form of $w_\star$

Given a dataset $\{x_i, y_i\}$ we can construct the optimal solution $w_\star$ to $\Phi_H$.

For convenience, we note that $\Phi_H$ can be rewritten as a minimization problem over a matrix $X$ and vectory $y$ constructed from the dataset as:

$$\boldsymbol{X}_i \equiv x_i \quad \boldsymbol{y}_i \equiv y_i \tag{1}$$

Then we have (using einstein convention for summing over repeated indices):

$$\begin{aligned}
\Phi &= \frac{1}{2n}\left(||\boldsymbol{y} - \boldsymbol{X}w||^2 + \lambda\,||w||^2\right) \\
&= \frac{1}{2n}\left(\left(y_i - x_{ik}w_k\right)\left(y_i - x_{ij}w_j\right) + \lambda w_m w_m\right) \\
&= \frac{1}{2n}\left(y_i y_i - y_i x_{ij} w_j - y_i x_{ik} w_k + x_{ij} x_{ik} w_k w_j + \lambda w_m w_m\right) \\
&= \frac{1}{2n}\left(y_i y_i - 2 y_i x_{ik} w_k + x_{ij} x_{ik} w_k w_j + \lambda w_m w_m\right)
\end{aligned} \tag{2}$$

Note that in the last line we have relabeled a dummy index to combine the cross terms from the previous line.

$w_\star$ should satisfy $\frac{\partial \Phi}{\partial w_{\star i}} = 0$. We can evaluate this derivative explicitly:

$$\begin{aligned}
\frac{\partial \Phi}{\partial w_l} &= \frac{\partial}{\partial w_l}\frac{1}{2n}\left(y_i y_i - 2 y_i x_{ik} w_k + x_{ij} x_{ik} w_k w_j + \lambda w_m w_m\right) \\
&= \frac{1}{2n}\left(0 - 2 y_i x_{ik} \partial_l w_k + x_{ik} x_{ij} \partial_l\left(w_k w_j\right) + \lambda \partial_l\left(w_m w_m\right)\right) \\
&= \frac{1}{2n}\left(-2 y_i x_{ik} \delta_{kl} + x_{ik} x_{ij}\left(\delta_{lk} w_j + \delta_{lj} w_k\right) + \lambda 2 \delta_{ml} w_l\right) \\
&= \frac{1}{n}\left(-y_i x_{il} + \left(x_{il} x_{ij} + \lambda \delta_{jl}\right) w_j\right)
\end{aligned} \tag{3}$$

(Again, in the last name we have relabeled dummy indices for convenience to combine terms.) Setting 3 equal to 0 we can solve:

$$
\begin{aligned}
0 &= \frac{1}{n}\left(-y_i x_{il} + \left(x_{il} x_{ij} + \lambda \delta_{jl}\right) w_j\right) \\
y_i x_{il} &= \left(x_{il} x_{ij} + \lambda \delta_{jl}\right) w_j \\
\left(x_{il} x_{ij} + \lambda \delta_{jl}\right)^{-1} y_i x_{il} &= w_j
\end{aligned} \tag{4}
$$

In standard matrix notation the final line is:

$$
w = \left(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{X}^\top y \tag{5}
$$