DSC 140A - Discussion 03

Problem 1.

Consider the **absolute loss**:

$$L_{\text{abs}}(H(\vec{x}), y) = |H(\vec{x}) - y|.$$

For a linear prediction rule $H(\vec{x}; \vec{w}) = \text{Aug}(\vec{x}) \cdot \vec{w}$, this takes the form:

$$L_{\text{abs}}(\vec{w}, \vec{x}, y) = |\,\text{Aug}(\vec{x}) \cdot \vec{w} - y|.$$

Show that a subgradient of L_{abs} with respect to $\vec{w}$ is:

$$\begin{cases} \text{Aug}(\vec{x}), & \text{if Aug}(\vec{x}) \cdot \vec{w} - y > 0 \\ -\text{Aug}(\vec{x}), & \text{if Aug}(\vec{x}) \cdot \vec{w} - y < 0 \\ \vec{0}, & \text{if Aug}(\vec{x}) \cdot \vec{w} - y = 0 \end{cases}$$

Solution:

$$\begin{aligned} L_{\text{abs}}(\vec{w}, \vec{x}, y) &= |\,\text{Aug}(\vec{x}) \cdot \vec{w} - y| \\ &= \begin{cases} \text{Aug}(\vec{x}) \cdot \vec{w} - y, & \text{if Aug}(\vec{x}) \cdot \vec{w} - y > 0 \\ -\text{Aug}(\vec{x}) \cdot \vec{w} + y, & \text{if Aug}(\vec{x}) \cdot \vec{w} - y < 0 \\ 0, & \text{if Aug}(\vec{x}) \cdot \vec{w} - y = 0 \end{cases} \end{aligned}$$

This function is non-differentiable at the $\vec{w}$ where $\text{Aug}(\vec{x}) \cdot \vec{w} = y$ (the third case).

Looking at the definition of subgradient, we see a valid subgradient $\vec{g} \in \mathbb{R}^n$ at a point $\vec{x}_0 \in \mathbb{R}^n$ for the function $f : \mathbb{R}^n \to \mathbb{R}$ if and only if $\forall \vec{z} \in \mathbb{R}^n$

$$f(\vec{z}) \geq f(\vec{x}_0) + \vec{g}^T(\vec{z} - \vec{x}_0)$$

Using that definition we will check if $\vec{g} = \vec{0}$ is a valid subgradient at $\vec{w}_0$ where $\text{Aug}(\vec{x}) \cdot \vec{w}_0 = y$ for L_{abs}

$$\begin{aligned} L_{\text{abs}}(\vec{z}) &\geq L_{\text{abs}}(\vec{w}_0) + \vec{g}^T(\vec{z} - \vec{w}_0) \\ |\,\text{Aug}(\vec{x}) \cdot \vec{w} - y| &\geq \vec{g}^T(\vec{z} - \vec{w}_0) \\ |\,\text{Aug}(\vec{x}) \cdot \vec{w} - y| &\geq 0 \end{aligned}$$

Therefore $\vec{g} = \vec{0}$ is a valid subgradient at $\vec{w}_0$.

As $L_{\text{abs}}(\vec{w}, \vec{x}, y)$ is differentiable for all other $\vec{w}$, we will now calulate their gradients.

When $\text{Aug}(\vec{x}) \cdot \vec{w} - y > 0$, $L_{\text{abs}}(\vec{w}, \vec{x}, y) = \text{Aug}(\vec{x}) \cdot \vec{w} - y$ so $\frac{\partial L_{\text{abs}}}{\partial \vec{w}} = \text{Aug}(\vec{x})$.

When $\text{Aug}(\vec{x}) \cdot \vec{w} - y < 0$, $L_{\text{abs}}(\vec{w}, \vec{x}, y) = -\text{Aug}(\vec{x}) \cdot \vec{w} + y$ so $\frac{\partial L_{\text{abs}}}{\partial \vec{w}} = -\text{Aug}(\vec{x})$.

Therefore a subgradient of L_{abs} with respect to $\vec{w}$ is

$$\begin{cases} \text{Aug}(\vec{x}), & \text{if Aug}(\vec{x}) \cdot \vec{w} - y > 0 \\ -\text{Aug}(\vec{x}), & \text{if Aug}(\vec{x}) \cdot \vec{w} - y < 0 \\ \vec{0}, & \text{if Aug}(\vec{x}) \cdot \vec{w} - y = 0 \end{cases}$$

.

Problem 2.

Recall that the regularized least squares risk is

$$\tilde{R}(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}) - y_i)^2 + \lambda\|\vec{w}\|^2$$

Show that

$$\tilde{R}(\vec{w}) = \frac{1}{n}\left(\vec{w}^T\Phi^T\Phi\vec{w} - 2\vec{w}^T\Phi^T\vec{y} + \vec{y}^T\vec{y}\right) + \lambda\vec{w}^T\vec{w},$$

where Φ is the matrix whose ith row is $\vec{\phi}(\vec{x}^{(i)}$, and where $\vec{y} = (y_1, \ldots, y_n)^T$.

Solution:

$$\begin{aligned}
\tilde{R}(\vec{w}) &= \frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}) - y_i)^2 + \lambda\|\vec{w}\|^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}((\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}))^2 - 2y_i\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}) + y_i^2) + \lambda\|\vec{w}\|^2 \\
&= \frac{1}{n}(\sum_{i=1}^{n}(\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}))^2 - 2\sum_{i=1}^{n}y_i\vec{w}\cdot\vec{\phi}(\vec{x}^{(i)}) + \sum_{i=1}^{n}y_i^2) + \lambda\vec{w}^T\vec{w} \\
&= \frac{1}{n}\left(\vec{w}^T\Phi^T\Phi\vec{w} - 2\vec{w}^T\Phi^T\vec{y} + \vec{y}^T\vec{y}\right) + \lambda\vec{w}^T\vec{w}
\end{aligned}$$

Problem 3.

In lecture, we defined a kernel function to be a function k which computes the dot product of vectors after they are mapped to some high-dimensional space. The useful thing about kernel functions is that they allow us to compute these dot products without actually mapping vectors them to the high-dimensional space, which can be costly. In this problem, we will consider the the 2nd-order *polynomial kernel*, defined to be

$$k(\vec{x}, \vec{x}') = (1 + \vec{x}\cdot\vec{x}')^2.$$

Let $\vec{\phi}(\vec{x}) : \mathbb{R}^3 \to \mathbb{R}^{10}$ be the mapping:

$$\vec{\phi}(\vec{x}) = (1, x_1^2, x_2^2, x_3^2, \sqrt{2}\,x_1, \sqrt{2}\,x_2, \sqrt{2}\,x_3, \sqrt{2}\,x_1x_2, \sqrt{2}\,x_1x_3, \sqrt{2}\,x_2x_3)^T,$$

where x_1, x_2, x_3 are the components of the input vector, $\vec{x}$. That is, $\vec{\phi}$ is a feature map which maps a vector into a higher-dimensional space.

Show that $k(\vec{x}, \vec{y}) = \vec{\phi}(\vec{x})\cdot\vec{\phi}(\vec{y})$. That is, that k indeed computes the inner product of vectors in the higher-dimensional space.

Solution:

$$\begin{aligned}
k(\vec{x}, \vec{y}) &= (1 + x_1y_1 + x_2y_2 + x_3y_3)^2 \\
&= 1 + x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2 + 2x_1y_1 + 2x_2y_2 + 2x_3y_3 + 2x_1x_2y_2y_1 + 2x_1x_3y_3y_1 + 2x_2x_3y_2y_3
\end{aligned}$$

Now,

$$\vec{\phi}(\vec{x}) \cdot \vec{\phi}(\vec{y}) = (1, x_1^2, x_2^2, x_3^2, \sqrt{2}\,x_1, \sqrt{2}\,x_2, \sqrt{2}\,x_3, \sqrt{2}\,x_1x_2, \sqrt{2}\,x_1x_3, \sqrt{2}\,x_2x_3) \begin{pmatrix} 1 \\ y_1^2 \\ y_2^2 \\ y_3^2 \\ \sqrt{2}\,y_1 \\ \sqrt{2}\,y_2 \\ \sqrt{2}\,y_3 \\ \sqrt{2}\,y_1y_2 \\ \sqrt{2}\,y_1y_3 \\ \sqrt{2}\,y_2y_3 \end{pmatrix}$$

$$= 1 + x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2 + 2x_1y_1 + 2x_2y_2 + 2x_3y_3 + 2x_1x_2y_2y_1 + 2x_1x_3y_3y_1 + 2x_2x_3y_2y_3$$

Therefore, they are equivalent.