Zining Fan, Weixuan Tang

1. In Definition 1, if any solution from $W^*(\hat{c})$ may be used by the loss function, a prediction model would then be incentivized to always predict $\hat{c} = 0$.

   Proof: If any solution from $W^*(\hat{c})$ may be used, the loss function essentially becomes

   $$\min_{w \in W^*(\hat{c})} c^T w - z^*(c) \tag{1}$$

   If we predict $\hat{c} = 0$, the optimization problem becomes $z^*(0) := \min_w 0^T w = 0$, and each $w \in S$ is an optimal solution, $W^*(0) = S$. Therefore, the loss function would be

   $$\min_{w \in S} c^T w - z^*(c) = c^T w^*(c) - z^*(c) = 0 \tag{2}$$

   Since our goal is to minimize loss function, the prediction model will be incentivized to always predict $\hat{c} = 0$.

2. Try to prove hinge loss and logistic loss is consistent with 0-1 loss.

   0-1 loss: $L(\hat{y}, y) = I(\hat{y} \neq y)$
   hinge loss: $L(\hat{y}, y) = \max(0, 1 - \hat{y} \cdot y)$
   logistic loss: $L(\hat{y}, y) = \log\left(1 + e^{-\hat{y} \cdot y}\right)$

   (a) For Binary Classification

   For binary Classification problem, $y \in (-1, +1), \hat{y} \in (-1, +1)$. This case is a special case for SPO loss?????????? So,

   $$L_{0-1}(\hat{y}, y) = \begin{cases} 0 & \text{if } y\hat{y} = 1 \\ 1 & \text{if } y\hat{y} = -1 \end{cases}$$

   $$L_{hinge}(\hat{y}, y) = \begin{cases} 0 & \text{if } y\hat{y} = 1 \\ 2 & \text{if } y\hat{y} = -1 \end{cases}$$

   $$L_{logistic}(\hat{y}, y) = \begin{cases} log(1 + e^{-1}) & \text{if } y\hat{y} = 1 \\ log(1 + e) & \text{if } y\hat{y} = -1 \end{cases}$$

   From the formulas above, we could see that:
   If we want to minimize hinge loss, we will set $y\hat{y} = 1$, in which case the 0-1 loss is also be minimized.
   Similarly, if we want to minimize Logistic loss, we will set $y\hat{y} = 1$, in which case the 0-1 loss is also be minimized.

   (b) For Regression

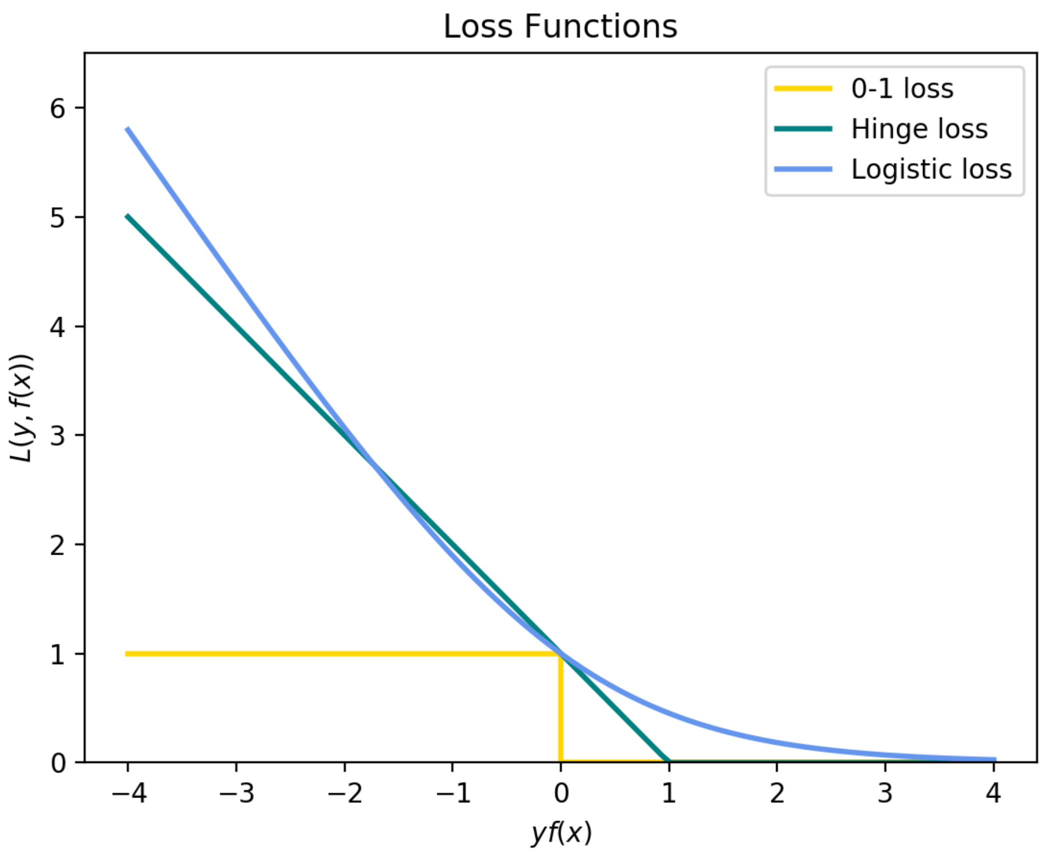
   For Regression problem, $y \in R, \hat{y} \in R$. So,

   $$L_{0-1}(\hat{y}, y) = \begin{cases} 0 & \text{if } y\hat{y} > 0 \\ 1 & \text{if } y\hat{y} <= 0 \end{cases}$$

$$L_{hinge}(\hat{y}, y) = \begin{cases} 0 & \text{if } y\hat{y} > 1 \\ 1 - y\hat{y} & \text{if } y\hat{y} <= 1 \end{cases}$$

$$L_{logistic}(\hat{y}, y) = log(1 + e^{-y\hat{y}})$$

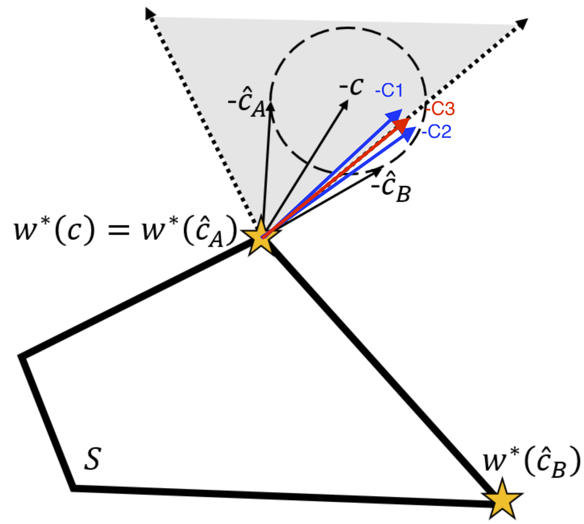
If we want to minimize hinge loss, we will set $y\hat{y} > 1$, in which case the 0-1 loss is also be minimized.
If we want to minimize Logistic loss, we will set $y\hat{y} = \infty$, in which case the 0-1 loss is also be minimized.



3. Prove that SPO loss is discontinuous and nonconvex.

   SPO loss: $\ell_{\text{SPO}}(\hat{c}, c) := \max_{w \in W^*(\hat{c})} \left\{ c^T w \right\} - z^*(c)$

   (a) Discontinuous



(a) Polyhedral feasible region

Take the figure on page 10 as an example. $\hat{c_1}$, $\hat{c_2}$ and $\hat{c_3}$ are three predicted values of c. $\hat{c_3}$ lies on the dotted line, and $\hat{c_1}$, $\hat{c_2}$ lie in different sides of the line. They result in different decisions, $w^*(\hat{c_1}) = w^*(\hat{c_3}) = w^*(\hat{c_A})$, $w^*(\hat{c_2}) = w^*(\hat{c_B})$. We consider the limit of $\ell_{\text{SPO}}(\hat{c}, c)$ at point $\hat{c_3}$. When we approach $\hat{c_3}$ from $\hat{c_1}$, the limit is $c^T w^*(\hat{c_A}) - z^*(c)$, but when we approach $\hat{c_3}$ from $\hat{c_2}$, the limit is $c^T w^*(\hat{c_B}) - z^*(c)$. The two limits are different, so the SPO loss is discontinuous at $\hat{c_3}$.

(b) Nonconvex

I would like to show that 0-1 loss for binary classification is non-convex.

0-1 loss function L is:

$$L(x) = \begin{cases} 0, & \text{if } x \geq 0 \\ 1, & \text{if } x < 0 \end{cases}$$

If L is convex, for $x_1, x_2 \in R$ and $\lambda \in [0,1]$, then:

$$L\left(\lambda x_1 + (1-\lambda)x_2\right) \leq \lambda L\left(x_1\right) + (1-\lambda)L\left(x_2\right)$$

Here is a counterexample, if $(x_1, x_2, \lambda) = (-1, 1/2, 1/2)$,

$$L\left(\lambda x_1 + (1-\lambda)x_2\right) = L(-1/4) = 1$$
$$L\left(x_1\right) = 1$$
$$L\left(x_2\right) = 0$$

So, $L\left(\lambda x_1 + (1-\lambda)x_2\right) \leq \lambda L\left(x_1\right) + (1-\lambda)L\left(x_2\right)$ does not hold. So, L is not convex.