P9120- Homework # 2

Assigned: October 10th, 2024 — Due at 9pm EST on October 26, 2024

Maximum points that you can score in this Homework is 20.
Please include R or python code you used to complete this homework.

1. (5 points) Let $X \in \mathbb{R}^p$ be the vector of input variables and $Y \in \{-1, 1\}$ be the binary outcome. Consider classification rule $G(X) = \text{sign}(f(X))$, where f is a real-valued function on $\mathbb{R}^p$. For loss functions $L(y, f(\mathbf{x}))$ given in (a)-(d) below, let $f^* = \arg\min_f EL(Y, f(X))$, where the expectation is taken over the joint distribution of X and Y. Show that

   (a) (Binomial Deviance loss) If $L(y, f(\mathbf{x})) = \log[1+\exp(-yf(\mathbf{x}))]$, then $f^*(\mathbf{x}) = \log \frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

   (b) (Hinge loss) If $L(y, f(\mathbf{x})) = [1-yf(\mathbf{x})]_+$, then $f^*(\mathbf{x}) = \text{sign}[Pr(Y = 1|X = \mathbf{x}) - \frac{1}{2}]$.

   (c) (quadratic loss) If $L(y, f(\mathbf{x})) = [y - f(\mathbf{x})]^2 = [1 - yf(\mathbf{x})]^2$, then $f^*(\mathbf{x}) = 2Pr(Y = 1|X = \mathbf{x}) - 1$.

   (d) (Exponential loss) If $L(y, f(\mathbf{x})) = \exp[-yf(\mathbf{x})]$, then $f^*(\mathbf{x}) = \frac{1}{2} \log \frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

2. (5 points) In this problem, we review *Lagrangian duality* discussed in Lecture 4 with a simple example. Consider the optimization problem.

$$\min_{x \in \mathbb{R}} x^2 - x + 1 \text{ subject to } x \geq 1. \tag{1}$$

   (a) Plot the objective function $f(x) = x^2 - x + 1$ and indicate on your plot the feasible set (i.e. the set of x that satisfies the constraint). From this plot, determine the minimizer x^* and the value of objective function at the minimizer $f(x^*)$.

   (b) Write down the Lagrangian primal function $L(x, \lambda)$, where λ is the Lagrangian multiplier. Derive and plot the Lagrangian dual function $L_D(\lambda)$.

   (c) State and solve the Lagrangian dual problem (by hand). Denote the maximizer by λ^* and the corresponding Lagrangian dual function by $L_D(\lambda^*)$. Does strong duality hold? (i,.e. does $L_D(\lambda^*) = f^(x^*)$?)

   (d) Based on part (b) and part (c), we can solve the optimization problem (1) by choosing x that minimizes $L(x, \lambda^*)$. Verify that the minimizer of $L(x, \lambda^*)$ is indeed x^*.

3. (5 points) Python Lab 1 exercise. (P9120_Lab1_Exercise.ipynb)

4. (5 points) Python Lab 2 exercise. (P9120_Lab2_Exercise.ipynb)

1. (5 points) Let $X \in \mathbb{R}^p$ be the vector of input variables and $Y \in \{-1, 1\}$ be the binary outcome. Consider classification rule $G(X) = \text{sign}(f(X))$, where f is a real-valued function on $\mathbb{R}^p$. For loss functions $L(y, f(\mathbf{x}))$ given in (a)-(d) below, let $f^* = \arg\min_f EL(Y, f(X))$, where the expectation is taken over the joint distribution of X and Y. Show that

(a) (Binomial Deviance loss) If $L(y, f(\mathbf{x})) = \log[1+\exp(-yf(\mathbf{x}))]$, then $f^*(\mathbf{x}) = \log \frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

(a) $f^* = \text{argmin}_f EL(Y, f(X)) = \text{argmin}_f E[\log(1+\exp(-Yf(X)))]$

$= \text{argmin}_f E[\log(1+\exp(-Yf(x))) \mid X=x]$

$= \text{argmin}_f (P(Y=1|X=x)\log(1+\exp(-f(x))) + P(Y=-1|X=x)\log(1+\exp(f(x)))$

$\frac{d}{df^*}(P(Y=1|X=x)\log(1+\exp(-f(x))) + P(Y=-1|X=x)\log(1+\exp(f(x))) = 0$

$\therefore P(Y=1|X=x)\frac{-\exp(-f(x))}{1+\exp(-f(x))} + P(Y=-1|X=x)\frac{\exp(f(x))}{1+\exp(f(x))} = 0$

$P(Y=1|X=x) = P(Y=-1|X=x)\exp(f(x))$

$\exp(f(x)) = \frac{P(Y=1|X=x)}{P(Y=-1|X=x)}$

$f(x)^* = \log\left(\frac{P(Y=1|X=x)}{P(Y=-1|X=x)}\right)$

(b) (Hinge loss) If $L(y, f(\mathbf{x})) = [1 - yf(\mathbf{x})]_+$, then $f^*(\mathbf{x}) = \text{sign}[Pr(Y=1|X=\mathbf{x}) - \frac{1}{2}]$.

HL=0 when $yf(x) \geq 1$
HL↑ linear $yf(x) < 1$

(b) $E[L(Y, f(X))] = P(Y=1|X=x)\max(0, 1-f(x)) + P(Y=-1|X=x)\max(0, 1+f(x))$

① $f(x) \geq 1$: $E[L(y, f(x))] = 0$

② $-1 \leq f(x) \leq 1$: $E[L(y, f(x))] = P(Y=1|X=x)(1-f(x)) + P(Y=-1|X=x)(1+f(x))$

$\frac{d}{df}E[L(Y, f(X))] = -P(Y=1|X=x) + P(Y=-1|X=x) = 0$

$\therefore P(Y=1|X=x) = P(Y=-1|X=-x) = \frac{1}{2}$

$f(x)^* = \text{sign}[P(Y=1|X=x) - \frac{1}{2}]$

(c) (quadratic loss) If $L(y, f(\mathbf{x})) = [y - f(\mathbf{x})]^2 = [1 - yf(\mathbf{x})]^2$, then $f^*(\mathbf{x}) = 2Pr(Y = 1|X = \mathbf{x}) - 1$.

(c) $E[L(Y, f(X))] = -2P(Y=1|X=x)(1-f(x)) + 2P(Y=-1|X=x)(-1-f(x)) = 0$

$P(Y=1|X=x)(1-f(x)) = P(Y=-1|X=x)(1+f(x))$

$P(Y=1|X=x) - P(Y=1|X=x)f(x) = P(Y=-1|X=x) + P(Y=-1|X=x)f(x)$

$\underbrace{(P(Y=-1|X=x) + P(Y=1|X=x))}_{=1} f(x) = P(Y=1|X=x) - P(Y=-1|X=x)$

$f(x)^* = 2P(Y=1|X=x) - 1$

(d) (Exponential loss) If $L(y, f(\mathbf{x})) = \exp[-yf(\mathbf{x})]$, then $f^*(\mathbf{x}) = \frac{1}{2}\log\frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

(d) $E[L(y, f(x))] = P(Y=1|X=x)\exp(-f(x)) + P(Y=-1|X=x)\exp(f(x))$

$\frac{d}{df} E[L(y, f(x))] = -P(Y=1|X=x)\exp(-f(x)) + P(Y=-1|X=x)\exp(f(x)) = 0$

$P(Y=1|X=x)\exp(-f(x)) = P(Y=-1|X=x)\exp(f(x))$

$P(Y=1|X=x) = P(Y=-1|X=x)\exp(2f(x))$

$\exp(2f(x)) = \frac{P(Y=1|X=x)}{P(Y=-1|X=x)}$

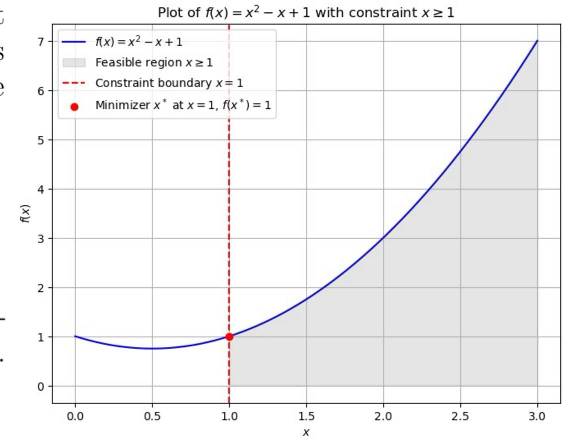
$2f(x) = \log\left(\frac{P(Y=1|X=x)}{P(Y=-1|X=x)}\right)$

$f(x)^* = \frac{1}{2}\log\left(\frac{P(Y=1|X=x)}{P(Y=-1|X=x)}\right)$

2. (5 points) In this problem, we review *Lagrangian duality* discussed in Lecture 4 with a simple example. Consider the optimization problem.

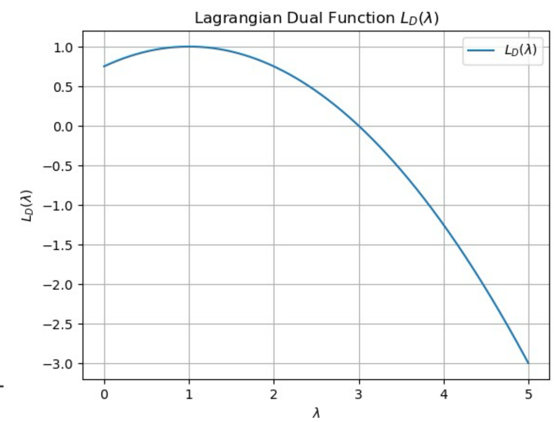
$$\min_{x\in\mathbb{R}} x^2 - x + 1 \text{ subject to } x \geq 1. \tag{1}$$

(a) Plot the objective function $f(x) = x^2 - x + 1$ and indicate on your plot the feasible set (i.e. the set of x that satisfies the constraint). From this plot, determine the minimizer x^* and the value of objective function at the minimizer $f(x^*)$.



On Jupyter notebook. $x^*=1$ $f(x^*)=1$

(b) Write down the Lagrangian primal function $L(x, \lambda)$, where λ is the Lagrangian multiplier. Derive and plot the Lagrangian dual function $L_D(\lambda)$.



$L(x,\lambda) = x^2 - x + 1 + \lambda(1-x)$

$\frac{dL(x,\lambda)}{dx} = 2x - 1 - \lambda = 0 \Rightarrow x = \frac{1+\lambda}{2}$

$L_D(\lambda) = L(\frac{1+\lambda}{2}, \lambda)$

$= (\frac{1+\lambda}{2})^2 - \frac{1+\lambda}{2} + 1 + \lambda(1 - \frac{1+\lambda}{2})$

Code on Jupyter notebook

(c) State and solve the Lagrangian dual problem (by hand). Denote the maximizer by λ^* and the corresponding Lagrangian dual function by $L_D(\lambda^*)$. Does strong duality hold? (i,.e. does $L_D(\lambda^*) = f(x^*)$?)

$L_D(\lambda) = \frac{(1+\lambda)^2}{4} - \frac{1+\lambda}{2} + 1 + \lambda(\frac{1-\lambda}{2})$

$= \frac{\lambda^2+2\lambda+1}{4} - \frac{1+\lambda}{2} + 1 + \frac{\lambda-\lambda^2}{2} = -\frac{\lambda^2}{4} + \frac{\lambda}{2} + \frac{3}{4}$

$\frac{dL_D(\lambda)}{dx} = -\frac{\lambda}{2} + \frac{1}{2} = 0 \Rightarrow \lambda^* = 1$

$L_D(\lambda^*) = -\frac{1}{4} + \frac{1}{2} + \frac{3}{4} = 1$

∴ It maximizes when $\lambda^* = 1$, then $L_D(\lambda^*) = 1 = f(x^*)$

Strong duality holds because the value of the dual function at λ^* equals the value of the primal function at x^*.

(d) Based on part (b) and part (c), we can solve the optimization problem (1) by choosing x that minimizes $L(x, \lambda^*)$. Verify that the minimizer of $L(x, \lambda^*)$ is indeed x^*.

$\lambda^* = 1 \Rightarrow L(x,\lambda^*) = x^2 - x + 1 + 1(1-x) = x^2 - 2x + 2$

$\frac{dL(x,\lambda^*)}{dx} = 2x - 2 = 0 \Rightarrow x^* = 1 = L_D(\lambda^*)$

∴ The minimizer of $L(x, \lambda^*)$ is indeed x^*.