NOTES. **NO** late submission will be accepted. Computer generated output without detailed explanations and remarks will not receive any credit. You may type out your answers, but make sure to use different fonts to distinguish your own words from computer output. Only hard copies are accepted. For the simulation and data analysis problems, keep the code you develop as you may be asked to present your work in class.

1. Consider the statement regrading (3.7) on P.81 of CO.

(a) Formulate the corresponding version precisely for concave functions. In particular, you should clearly specify the domains of all involved functions.

(b) Use Part (a) to show that the dual function is always concave.

2. Show whether strong duality holds for ridge regression and Lasso when $t = 0$.

3. Show that the set $\mathcal{A}$ defined in (5.37) on P.232 of CO is convex, if the underlying optimization problem is convex.

4. Exercise 5.1 (a), (b), (c) of CO.

5. Exercise 5.14 of CO.

6. Exercise 5.27 of CO, assuming that the primal problem is feasible, i.e. $Gx = h$ has solutions.

7. Assume $\boldsymbol{X} \in \mathbb{R}^{N\times p}$ is full rank, i.e. $\text{rank}(\boldsymbol{X}) = p$, and assume $\boldsymbol{y} \neq \boldsymbol{0}$. Consider the two formulations of Lasso. The first one is

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \\ \text{subject to} \quad & \|\boldsymbol{\beta}\|_1 \leq t. \end{aligned} \tag{P}$$

Denote by $\hat{\boldsymbol{\beta}}(t)$ the optimal solution of (P). The other formulation is

$$\min_{\boldsymbol{\beta}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\}. \tag{D}$$

Denote by $\hat{\boldsymbol{\beta}}_\lambda$ the optimal solution of (D). Let $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ be the LSE. Prove that for each $0 < t < \|\hat{\boldsymbol{\beta}}\|_1$, there exists a $\lambda > 0$, such that $\hat{\boldsymbol{\beta}}_\lambda = \hat{\boldsymbol{\beta}}(t)$.

Statistical Learning HW1 Applied

Yaniv Bronshtein

10/16/2021

(a). *Analysis of Primal Problem* Give the feasible set, the optimal value, and the optimal solution

Feasible Set: The interval [2,4] $(x-2)(x-4) \leq 0$ $x-2 \leq 0$ and $x-4 \leq 0$ $x \leq 2$ and $x \leq 4$ Thus, the optimal point is x*=2 The optimal value is $2^2+1=5$

(b). *Lagrangian and dual Function* Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal optimal point and value, and plot the Lagrangian $L(x,\lambda)$ versus x for a few positive values of λ. Verify the lower bound property $p* \geq inf L(x,\lambda)$ for $\lambda \geq 0$. Derive and sketch the Lagrange dual function g.

```
x <- seq(-5,5, 0.1)
f0 <- x^2+1
f1 <- (x-2)*(x-4)
par(mfrow=c(1,2))
plot(
  x=x,
  y=f0,
  main="f0 and f1",
  ylab="",
  xlab="x",
  type="l",
  ylim = c(-5,25),
  col="blue"
)
lines(x=x, y=f1, col="magenta")
legend(
  "topleft",
  c("f0=x^2+1","f1=(x-2)(x-4)"),
  fill=c("blue","magenta")
)

abline(v=2, col='red')
abline(v=4, col='red')
abline(h=0, col='black')
```

1.

a) If for each $y \in A$, $f(x,y)$ is concave in x then the function g defined as

$$g(x) = \inf_{y \in A} f(x,y)$$

is concave in x. Here the domain of g is

$$\text{dom } g = \{x \mid (x,y) \in \text{dom } f \text{ for all } y \in A, \inf_{y \in A} f(x,y) > -\infty\}$$

The pointwise infimum of a set of concave functions is a concave function.

b) The Lagrange function $L(x, \lambda, \nu)$ is affine in λ and ν. As such, it must be concave. Based on Part (a), the infimum of concave functions is concave.

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

2. Recall

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 \leq t$$

When $t=0$, $\sum_{j=1}^{p} \beta_j^2 \leq 0$

Since a square must be ≥ 0, it follows that $\sum_{j=1}^{p} \beta_j^2 = 0$

We re-write using the first definition

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\}$$

Thus we simplify to the Least Squares Method for fitting a regression.

Recall that Strong Duality is when $p^* = d^*$ which is when the optimal duality gap is equal to 0.

One way to achieve this is to show that Slater's Condition holds and that the problem is convex.

(1) The problem is indeed convex because we are seeking to minimize X^TX subject to the constraint $Ax = b$.

② In our case Slater's Condition is that the primal problem is feasible.
If $b \in R(A)$ $(p^* < \infty)$

For reference we show below how to derive the dual problem:
We first define the Lagrangian
$L(x, v) = x^T x + v^T (A^T x - b)$

The lagrangian dual is
$F(v) = \inf L(x, v) = L(-\frac{1}{2} v^T A, v) = \frac{1}{4} v^T A A^T v - \frac{1}{2} v^T A A^T v - v^T b$
$= -\frac{1}{4} v^T A A^T v - v^T b$

Let us now look at Lasso:
$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_i \beta_j)^2$

subject to $\sum_{j=1}^{p} |\beta_j| \leq t$

If $t = 0$, then $\sum_{j=1}^{p} |\beta_j| = 0$
We modify $\hat{\beta}^{lasso}$.
$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0)^2$

thus arriving at the same Least Squares Problem faced with Ridge

3. $A \subseteq G + (\mathbb{R}^m_+ \times \{0\} \times \mathbb{R}_+)$

or more explicitly,

$$A = \{(u, v, t) \mid \exists x \in D, f_i(x) \leq u_i, i = 1, \ldots, m, \\ h_i(x) = v_i, i = 1, \ldots, p \; f_0(x) \leq t\}$$

An optimization problem is convex if the bounds on the variables restrict the domain of the objective and constraints to a region where the functions are convex.

For $(u_1, v_1, t_1) \in A$ we can find x_1 such that

$f_i(x_1) \leq u_1$

$h_i(x_1) = v_1$

$f_0(x_1) \leq t_1$.

For $(u_2, v_2, t_2) \in A$ we can find x_2 such that

$f_i(x_2) \leq u_2$

$h_i(x_2) = v_2$

$f_0(x_2) \leq t_2$

We are given that the problem is convex.

Thus, we know that $f_i(x)$ is convex and $h_i(x)$ is linear.

We can write the following:

$f_i(\theta x_1 + (1-\theta) x_2) = \theta f_i(x_1) + (1-\theta) f_i(x_2) \leq \theta u_1 + (1-\theta) u_2$

$h_i(\theta x_1 + (1-\theta) x_2) = \theta h_i(x_1) + (1-\theta) h_i(x_2) = \theta v_1 + (1-\theta) v_2$

$f_0(x)$ is convex

Therefore,

$\theta(u_1, v_1, t_1) + (1-\theta)(u_2, v_2, t_2) \in A$

and A is convex.

4.

Exercise 5.1 a, b, c

5.1 A simple example. Consider the optimization problem
minimize x^2+1
subject to $(x-2)(x-4)\leq 0$
with variable $x \in \mathbb{R}$

a) Analysis of primal problem. Give the feasible set, the optimal value, and the optimal solution.

Feasible set: The interval $[2,4]$.

$(x-2)(x-4)\leq 0$

$x-2\leq 0 \quad x-4\leq 0$

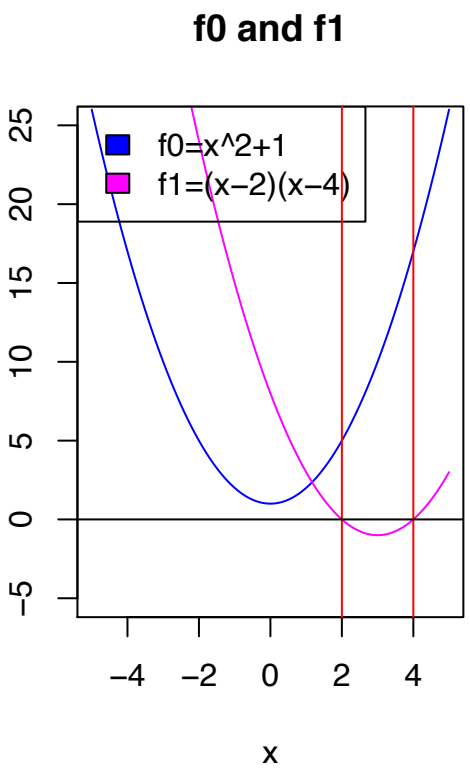
$x\leq 2 \quad x\leq 4$

Thus, the optimal point is $x^*=2$.

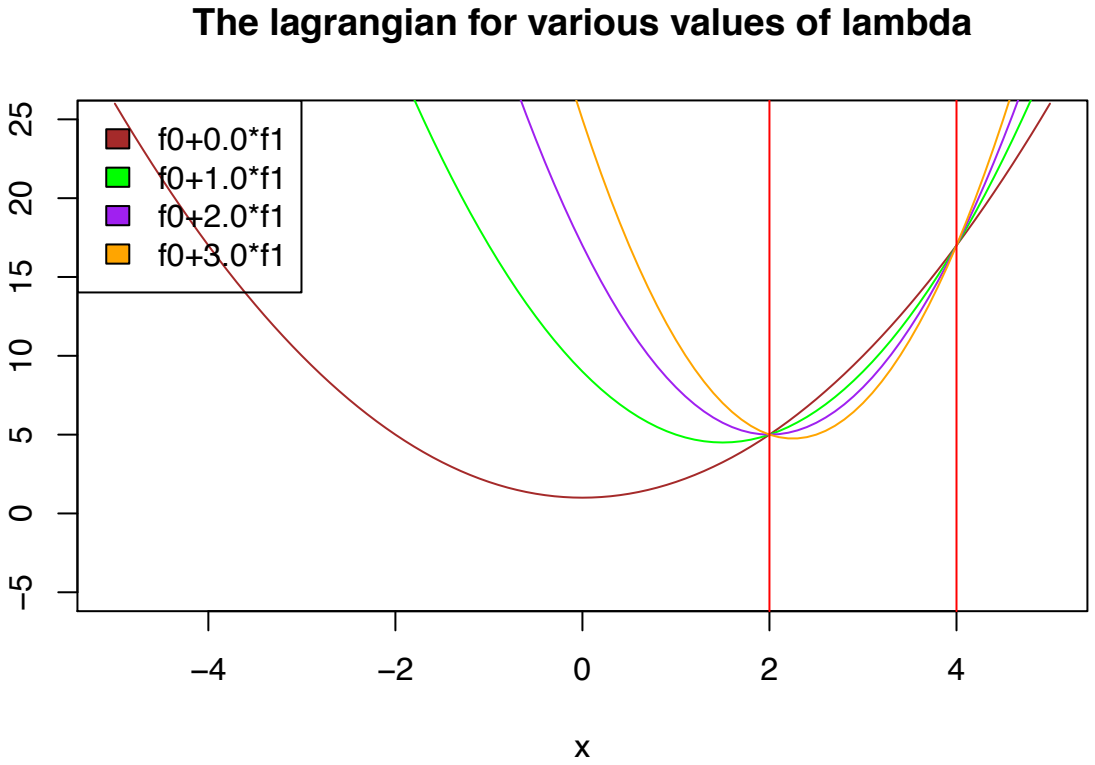
$2^2+1=5$

so the optimal value is $p^*=5$

b) Lagrangian and dual function. Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian $L(x,\lambda)$ versus x for a few positive values of λ. Verify the lower bound property ($p^* \geq \inf_x L(x,\lambda)$ for $\lambda \geq 0$). Derive and sketch Lagrange dual

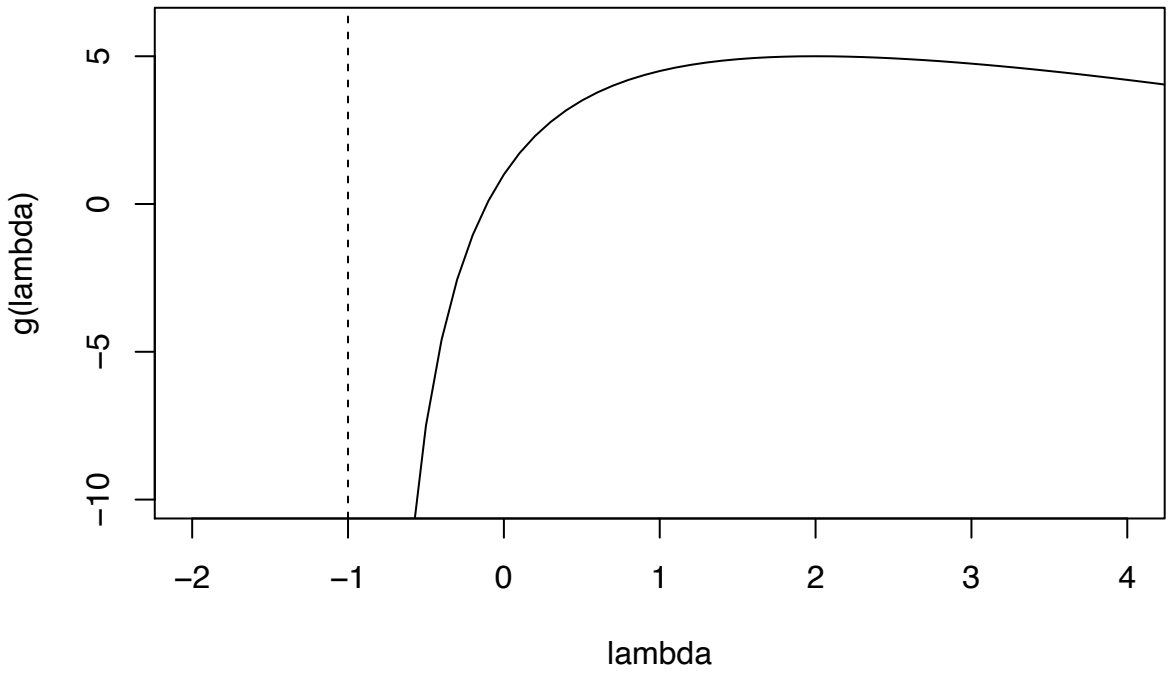



```
plot(x=x,
     y=f0,
     ylab="",
     xlab="x",
     type="l",
     ylim = c(-5,25),
     col="brown",
     main="The lagrangian for various values of lambda"
)
lines(x,f0+1.0*f1, col="green")
lines(x,f0+2.0*f1, col="purple")
lines(x,f0+3.0*f1, col="orange")

legend(
  "topleft",
  c("f0+0.0*f1","f0+1.0*f1", "f0+2.0*f1", "f0+3.0*f1"),
  fill=c("brown","green", "purple", "orange")
)

abline(v=2, col='red')
abline(v=4, col='red')
```



The overlayed plot above demonstrates the Lagrangian with input x and λ as the sum of f_0 and f_1 times a constant λ The minimum value of the Lagrangian is always less than p*. The maximum is reached at a λ value of 2 and decreases after that.

```
lambda <- seq(-0.9,16/3,0.1);
g <- (-9*lambda^2)/(1+lambda) + 1 + 8*lambda
plot(x=lambda,
     y=g,
     ylab="g(lambda)",
     xlab="lambda",
     type="l",
     xlim=c(-2,4),
     ylim=c(-10,6)
)
abline(v=-1, lty='dashed')
```


g(lambda)
5
0
−5
−10
−2
−1
0
1
2
3
4
lambda

For $\lambda > -1$, the Lagrangian reaches its minimum at $u = \frac{3\lambda}{1+\lambda}$

For $\lambda \leq -1$ it is unbounded

Thus $g(\lambda) = \begin{cases} -9\frac{\lambda^2}{(1+\lambda)} + 1 + 8\lambda & \lambda > -1 \\ -\infty & \lambda \leq -1 \end{cases}$

The dual function is concave and $p^* = 5$ for $\lambda = 2$ and less for other values.

c) Lagrange Dual Problem. State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution λ^*. Does Strong Duality hold?

In our case, the dual problem is to

Maximize $\frac{-9\lambda^2}{1+\lambda} + 1 + 8\lambda$

Constraint: $\lambda \geq 0$

The dual optimum is at $\lambda = 2$ with $d^* = 5$.

Slater's constraint qualifications are satisfied.

5. A penalty method for equality constraints.

We consider the problem

minimize $f_0(x)$

subject to $Ax=b$

where $f_0: \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, and $A \in \mathbb{R}^{m \times n}$ with rank $A=m$.

In a quadratic penalty method, we form an auxiliary function

$\phi(x) = f_0(x) + \alpha \|Ax-b\|_2^2$ where $\alpha > 0$ is a param.

:

show how to find, from $\tilde{x}$, a dual feasible point.

If $\tilde{x}$ minimizes ϕ then we can write the following:

$\nabla f_0(\tilde{x}) + 2\alpha A^T(A\tilde{x}-b) = 0$

We observe that $\tilde{x}$ is also a minimizer of

$\nabla f_0(x) + v^T(Ax-b)$

where $v = 2\alpha(A\tilde{x}-b)$.

v is dual feasible with $g(v)$:

$g(v) = \inf_x (f_0(x) + v^T(Ax-b)) = f_0(\tilde{x}) + 2\alpha\|A\tilde{x}-b\|_2^2$

$f_0(x) \geq f_0(\tilde{x}) + 2\alpha\|A\tilde{x}-b\|_2^2$ $\forall x$ that satisfy $Ax=b$

6. Equality Constrained Least-Squares.

Consider the equality constrained least squares problem

$$\text{minimize } \|Ax-b\|_2^2$$
$$\text{subject to } Gx=h$$

where $A \in \mathbb{R}^{m \times n}$ with rank $A=n$, and $G \in \mathbb{R}^{p \times n}$ with rank $G=p$

Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution v^*.

We also assume that the primal problem is feasible meaning that $Gx=h$ has solutions

Let us first define the Lagrangian:

$$L(x,v) = \|Ax-b\|_2^2 + v^T(Gx-h)$$
$$= x^TA^TAx + (G^Tv-2A^Tb)^Tx - v^Th$$

Our minimizer:

$$x = -\frac{1}{2}(A^TA)^{-1}(G^Tv-2A^Tb)$$

We now formulate the dual function:

$$g(v) = -\frac{1}{4}(G^Tv-2A^Tb)^T(A^TA)^{-1}(G^Tv-2A^Tb) - v^Th$$

Our two optimality conditions:

① $2A^T(Ax^* - b) + G^T v^* = 0$

② $Gx^* = h$

Let us solve for x^* in ①

$2A^TAx^* - 2A^Tb + G^Tv^* = 0$

$2A^TAx^* = 2A^Tb - G^Tv^*$

$$x^* = (A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^Tv^*\right)$$

Let us plug x^* into ② to derive for v^*.

$G\left((A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^Tv^*\right)\right) = h$

$G(A^TA)^{-1}A^Tb - G(A^TA)^{-1}\left(\frac{1}{2}G^Tv^*\right) = h$

$G(A^TA)^{-1}\left(\frac{1}{2}G^Tv^*\right) = h - G(A^TA)^{-1}A^Tb$

$$v^* = -2\left(G(A^TA)^{-1}G^T\right)^{-1}\left(h - G(A^TA)^{-1}A^Tb\right)$$

7. We know that $f_0(x) = \|y - X\beta\|^2$ is convex.
$f_1(x) = \|\beta\|_1 - t$ is also convex.
Thus, Slater's condition is applicable for the following:
① $\|\beta\|_1 \leq t$
② $0 < t < \|\hat{\beta}\|$

Strong Duality holds.
We define the Lagrangian as follows:
$L(\beta, \lambda) = \|y - X\beta\|^2 + \lambda^*(\|\beta\|_1 - t)$. for $\lambda > 0$

To stick to the notation:
$d^* = \min_{\beta}(\|y - X\beta\|^2 + \lambda^*(\|\beta\|_1 - t)) = g(\lambda^*)$
Distributing λ^*:
$d^* = \min_{\beta}(\|y - X\beta\|^2 + \lambda^*\|\beta\|_1 - \lambda^* t)$

This in fact equal to $\hat{\beta}_{t_1}$.

β_{λ^*} would simply be $\min_{\beta}(\|y - X\beta\|^2 + \lambda^*\|\beta_1\|_1)$

KKT holds ($\hat{\beta}_t = \hat{\beta}_{\lambda^*}$)