Solutions for Home Assignment №3

November 7, 2024

Exercise 1

[3 points]. Prove the following matrix identity

$$(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} = PB^T (BPB^T + R)^{-1}, \tag{1}$$

where $P \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, and $R \in \mathbb{R}^{m\times m}$. P and R are invertible. Note that if $m \ll n$, it will be much cheaper to evaluate the right-hand side than the left-hand side. Hint: right multiplying both sides by $(BPB^T + R)$. With similar arguments, prove a special case of Eq. (1)

$$(I + AB)^{-1} A = A(I + BA)^{-1},$$

where $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times n}$.

Solution:

Proof. We manipulate the above matrix identity by right multiplying both sides by $(BPB^T + R)$. For the left-hand side, we have

$$\begin{aligned}
&(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} (BPB^T + R) \\
=&(P^{-1} + B^T R^{-1} B)^{-1} (B^T R^{-1} BPB^T + B^T R^{-1} R) \\
=&(P^{-1} + B^T R^{-1} B)^{-1} (B^T R^{-1} BPB^T + P^{-1} PB^T) \\
=&(P^{-1} + B^T R^{-1} B)^{-1} (B^T R^{-1} B + P^{-1}) PB^T \\
=&PB^T.
\end{aligned}$$

It is trivial that the right-hand side is also equal to PB^T. For the special case, we have

$$
\begin{aligned}
&(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} \\
=&(P^{-1}(PP^{-1} + PB^T R^{-1} B))^{-1} B^T R^{-1} \\
=&(I + \underbrace{PB^T R^{-1}}_{A} \underbrace{B}_{B})^{-1} \underbrace{PB^T R^{-1}}_{A},
\end{aligned}
$$

and

$$
\begin{aligned}
&PB^T (BPB^T + R)^{-1} \\
=&PB^T ((BPB^T R^{-1} + RR^{-1})R)^{-1} \\
=&\underbrace{PB^T R^{-1}}_{A} (\underbrace{B}_{B} \underbrace{PB^T R^{-1}}_{A} + I)^{-1}.
\end{aligned}
$$

Hence, assuming $A = PB^T R^{-1}$ and $B = B$ (as illustrated above), we obtain

$$(I + AB)^{-1} A = A(I + BA)^{-1}.$$

$\square$

Exercise 2

[5 points]. Say you have M linear equations in N variables. In matrix form we write $Ax = y$, where $A \in \mathbb{R}^{M \times N}$, $x \in \mathbb{R}^{N \times 1}$, and $y \in \mathbb{R}^{M \times 1}$. Given a proof or a counterexample for each of the following.

a) [1 point]. If $N = M$, there is always *at most one solution*.

b) [1 point]. If $N > M$, you can *always* solve $Ax = y$.

c) [1 point]. If $N > M$, the nullspace of A has dimension greater than zero.

d) [1 point]. If $N < M$, then for *some* y there is *no* solution of $Ax = y$.

e) [1 point]. If $N < M$, the *only* solution of $Ax = 0$ is $x = 0$.

Hint: The null space of A, denoted by V, contains the set of vectors that satisfy $\{x \in V | Ax = 0\}$.

Solution:

a) False. One counterexample is A=$\begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}$, y=$\begin{bmatrix} 4 \\ 8 \end{bmatrix}$.

b) False. One counterexample is A=$\begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix}$, y=$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

c) True.

Proof. From the Rank-nullity theorem we have

$$\text{Rank}(A) + \text{Nullity}(A) = N.$$

We also know that

$$\text{Rank}(A) \leq M.$$

Thus

$$\text{Nullity}(A) = N - \text{Rank}(A) \geq N - M > 0.$$

$\square$

d) True. One example is A=$\begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 2 & 2 \end{bmatrix}$, y=$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

e) False. One counterexample is A=$\begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix}$.

Exercise 3

[4 points]. Coordinate Descent for Linear Regression. We would like to solve the following linear regression problem

$$\text{minimize} \sum_{i=1}^{M} (y^{(i)} - w^T x^{(i)})^2, \tag{2}$$

where $w \in \mathbb{R}^{N \times 1}$ and $x^{(i)} \in \mathbb{R}^{N \times 1}$ using coordinate descent.

a) [2 points]. In the current iteration, w_k is selected for update. Please prove the following update rule:

$$w_k \leftarrow \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})}{\sum_{i=1}^{M} (x_k^{(i)})^2}, \quad \forall k \in \{1, 2, \dots, N\} \tag{3}$$

b) [2 points]. Prove that the following update rule for w_k is equivalent to Eq. (3).

$$w_k^{\text{old}} \leftarrow w_k, \tag{4}$$

$$w_k \leftarrow \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot r^{(i)}}{\sum_{i=1}^{M} (x_k^{(i)})^2} + w_k^{\text{old}}, \tag{5}$$

$$r^{(i)} \leftarrow r^{(i)} + (w_k^{\text{old}} - w_k) x_k^{(i)} \quad \forall i \in \{1, 2, \dots M\}. \tag{6}$$

where $r^{(i)}$ is the residual

$$r^{(i)} = y^{(i)} - \sum_{j=1}^{N} w_j x_j^{(i)}. \tag{7}$$

Compare the two update rules. Which one is better and why?

Solution:

a) *Proof.*

$$f(w) = \sum_{i=1}^{M} (y^{(i)} - w^T x^{(i)})^2, \tag{8}$$

Find a closed form solution:

$$\frac{\partial f(w)}{\partial w_k} = 0 \tag{9}$$

So:

$$\frac{\partial f(w)}{\partial w_k} = \sum_{i=1}^{M} 2(-x_k^{(i)}) \cdot (y^{(i)} - \sum_{j=1}^{N} w_j x_j^{(i)}) = 0 \tag{10}$$

$$\sum_{i=1}^{M}(y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)} - w_k x_k^{(i)})x_k^{(i)} = 0 \tag{11}$$

$$\sum_{i=1}^{M}(y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})x_k^{(i)} - \sum_{i=1}^{M} w_k (x_k^{(i)})^2 = 0 \tag{12}$$

$$w_k \sum_{i=1}^{M}(x_k^{(i)})^2 = \sum_{i=1}^{M}(y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})x_k^{(i)} \tag{13}$$

$$w_k \leftarrow \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})}{\sum_{i=1}^{M}(x_k^{(i)})^2}, \quad \forall k \in \{1, 2, \dots, N\} \tag{14}$$

□

b) *Proof.* Rewrite the expressions above as

$$w_{k(t+1)}^{\text{old}} = w_{k(t)}, \tag{15}$$

$$w_{k(t+1)} = \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot r_{(t)}^{(i)}}{\sum_{i=1}^{M}(x_k^{(i)})^2} + w_{k(t+1)}^{\text{old}}, \tag{16}$$

$$r_{(t+1)}^{(i)} = r_{(t)}^{(i)} + (w_{k(t+1)}^{\text{old}} - w_{k(t+1)})x_k^{(i)}, \tag{17}$$

$$r_{(t)}^{(i)} = y^{(i)} - \sum_{j=1}^{N} w_{j(t)} x_j^{(i)} = y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)} - w_{k(t)} x_k^{(i)} \tag{18}$$

Thus we have

$$
\begin{aligned}
w_{k(t+1)} &= \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)} - w_{k(t)} x_k^{(i)})}{\sum_{i=1}^{M} (x_k^{(i)})^2} + w_{k(t)} \\
&= \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)}) - \sum_{i=1}^{M} w_{k(t)} (x_k^{(i)})^2}{\sum_{i=1}^{M} (x_k^{(i)})^2} + w_{k(t)} \\
&= \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})}{\sum_{i=1}^{M} (x_k^{(i)})^2} - w_{k(t)} + w_{k(t)} \\
&= \frac{\sum_{i=1}^{M} x_k^{(i)} \cdot (y^{(i)} - \sum_{j=1, j\neq k}^{N} w_j x_j^{(i)})}{\sum_{i=1}^{M} (x_k^{(i)})^2}.
\end{aligned}
$$

□

The latter is better. Because the cost for the former is $O(m \cdot n^2)$, but the cost for the latter one is $O(m \cdot n)$.

Exercise 4

[3 points]. Consider the soft-margin SVM problem using an ℓ_2-norm penalty on the slack variables,

$$
\begin{aligned}
&\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i^2 \\
&\text{s.t. } y_i \left(w^T x_i + b \right) \geq 1 - \xi_i, \quad \forall i \\
&\quad \xi_i \geq 0, \quad \forall i,
\end{aligned} \tag{19}
$$

where ξ_i is the slack variable that allows the i th point to violate the margin.

a) [1 point]. Show that the non-negative constraint on ξ_i is redundant, and hence can be dropped. Hint: show that if $\xi_i < 0$ and the margin constraint is satisfied, then $\xi_i = 0$ is also a solution with lower cost.

b) [1 point]. Derive the Lagrangian.

c) [1 point]. Derive the SVM dual problem.

Solution:

a) If $\xi_i < 0$ and the constraint $y_i\left(w^T x_i + b\right) \geq 1 - \xi_i$ is satisfied, then the constraint is satisfied by $\xi_i = 0$ with lower cost. Hence $\xi_i < 0$ is never a solution.

b) As the constraint on ξ_i can be dropped, the Lagrangian is

$$L(w, b, \xi, \alpha) = \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right), \quad (20)$$

where α_i are Lagrange multipliers.

c) Take the partial derivatives of L w.r.t. w, b, ξ_i and set them to zero

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0, \quad (21)$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i y_i = 0, \quad (22)$$

$$\frac{\partial L}{\partial \xi_i} = C\xi_i - \alpha_i = 0. \quad (23)$$

We have

$$w^* = \sum_{i=1}^{n}\alpha_i y_i x_i, \quad (24)$$

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \quad (25)$$

$$\xi_i^* = \frac{\alpha_i}{C}. \quad (26)$$

Plugging the above equations into the Lagrangian, we have

$$\begin{aligned} L(\alpha) &= \frac{1}{2}(\sum_{i=1}^{n}\alpha_i y_i x_i)^2 + \frac{C}{2}\sum_{i=1}^{n}\frac{\alpha_i^2}{C^2} - \sum_{i=1}^{n}\alpha_i y_i(\sum_{j=1}^{n}\alpha_j y_j x_j^T)x_i - b\sum_{i=1}^{n}\alpha_i y_i + \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i * \frac{\alpha_i}{C} \\ &= \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{1}{2}\sum_{i=1}^{n}\frac{\alpha_i^2}{C} \\ &= \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i^T x_j + \frac{1}{C}\delta_{ij}\right), \end{aligned} \quad (27)$$

where $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. Hence, the SVM dual problem is

$$\begin{aligned} \max_{\alpha} \ & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(x_i^T x_j + \frac{1}{C} \delta_{ij} \right) \\ \text{s.t.} \ & \sum_{i=1}^{n} \alpha_i y_i = 0, \\ & \alpha_i \geq 0, \quad \forall i. \end{aligned} \tag{28}$$