Home Assignment №3

Due on November 8, 2024, 11:59pm

Exercise 1

[3 points]. Prove the following matrix identity

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

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. Using Eq. (1), prove a special case

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

where $A \in \mathbb{R}^{N\times M}$ and $B \in \mathbb{R}^{M\times N}$.

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 $x = 0$.

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

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:

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

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

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

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

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

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?

Exercise 4

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

$$
\begin{aligned}
&\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i^2 \\
&\text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 - \xi_i, \quad \forall i \\
&\quad \xi_i \geq 0, \quad \forall i,
\end{aligned} \tag{8}
$$

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.