Recitation Week 12

Ashwin Bhola

CDS, NYU

Nov 20th , 2019

Linear Regression

1. When solving the least squares problem, the optimization problem is: $\min_x ||Ax - b||^2, A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n, x \in \mathbb{R}^d$
    1. Geometrically, what does this mean?
    2. How can we obtain the normal equation $A^T Ax = A^T b$ from this geometric intuition?
    3. Under what conditions is $A^T A$ invertible? If $A^T A$ is not invertible, must the normal equations still have a solution?
    4. The objective of linear regression is to find a hyperplane of the form $f(a_i) = x_0 + x^T a_i = x_0 + a_i^1 x_1 + \cdots a_i^d x_d$. Solving least squares gives us $x \in \mathbb{R}^d$. How do we find $x_0 \in \mathbb{R}$?
    5. One of the things we expect from machine learning models is regularity in prediction i.e. neighboring points in feature space should have similar prediction. How does regularization induce this regularity?
    6. In ridge regression, we add a penalty $\lambda||x||_2^2$. Is it better than adding $\lambda||x||_2$?
    7. What are some downsides of using lasso as a feature selection tool?

Linear Regression

1. When solving the least squares problem, the optimization problem is: $\min_x ||Ax - b||^2, A \in \mathbb{R}^{n \times d}, b \in \mathbb{R}^n, x \in \mathbb{R}^d$
   1. This means that we are trying to find a point in $Im(A) \subset \mathbb{R}^n$ which is closest to $b \in \mathbb{R}^n$
   2. The point closest to b in the subspace $Im(A)$ is the projection of b onto $Im(A)$. Let $P_{Im(A)}b = Ax$

$$\Rightarrow Ax - b \perp Im(A)$$

$$\Rightarrow Ax - b \in Ker(A^T)$$

$$\Rightarrow A^T(Ax - b) = 0$$

$$\Rightarrow A^TAx = A^Tb$$

   3. A^TA is invertible if rank(A)=d. Yes, the equations always have a solution because $Im(A^TA) = \mathrm{Im}(\mathrm{A}^\mathrm{T})$
   4. Add a column of ones to A
   5. $f(m + \epsilon) - f(m) = x^T(m + \epsilon) - x^T(m) = x^T\epsilon \leq ||x||||\epsilon||$ where ϵ is just small noise
   6. $||x||$ is neither strictly convex nor differentiable at $x = 0$
   7. [Discussed in detail during the recitation]

Linear Regression

1. Let $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times 2}$ have the form

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$$

   Let $x \in \mathbb{R}^2$ be the minimizer of the least squares objective function $\min_x ||Ax - b||^2$. Prove that $\sum_{i=1}^{n} (Ax - b)_i = 0$

2. Suppose $M \in \mathbb{R}^{2\times 2}$ has positive eigenvalues. Describe geometrically the contour lines of $f: \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = x^T M x$. Recall that the contour line for value γ is given by $\{x \in \mathbb{R}^2 \mid f(x) = \gamma\}$

Linear Regression

1. Let $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times 2}$ have the form

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$$

Let $x \in \mathbb{R}^2$ be the minimizer of the least squares objective function $\min_x ||Ax - b||^2$. Prove that $\sum_{i=1}^{n} (Ax - b)_i = 0$

Solution: If x is the solution, then $A^T(Ax - b) = 0$. Let $v = Ax - b \Rightarrow A^T v = 0$

$$\Rightarrow 1^T v = 0$$

$$\Rightarrow \sum_{i=1}^{n} v_i = \sum_{i=1}^{n} (Ax - b)_i = 0$$

Linear Regression

2. Suppose the symmetric matrix $M \in \mathbb{R}^{2\times 2}$ has positive eigenvalues. Describe geometrically the contour lines of $f: \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = x^T M x$. Recall that the contour line for value γ is given by $\{x \in \mathbb{R}^2 \mid f(x) = \gamma\}$

Solution:

Since M is symmetric, $M = UDU^T$. Since U is orthogonal, the columns u_1, u_2 of U are the basis of $\mathbb{R}^2$. Let $x = x_1 u_1 + x_2 u_2$

Thus, $x^T M x = x_1^2 d_1 + x_2^2 d_2$ where $d_1, d_2 > 0$ are the diagonal entries of D.

$$\Rightarrow x_1^2 d_1 + x_2^2 d_2 = \gamma$$

$$\Rightarrow \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

Where $a^2 = \frac{\gamma}{d_1}$ and $b^2 = \frac{\gamma}{d_2}$. Thus, the contour lines take shape of an ellipse centered at 0, with the size of ellipse increasing as γ increases

Linear Regression

Suppose $A \in \mathbb{R}^{n\times 2}$ is your data matrix. The empirical risk or the least squares loss is given by $\mathrm{R}(\mathrm{w}) = \left||Aw - y|\right|^2$. Suppose

$\hat{w} = argmin_w\, R(w)$. Let $\mathrm{M} = A^T A \in \mathbb{R}^{2\times 2}$ be invertible. It can be easily verified that

$$x^T M x - 2b^T x = (x - M^{-1}b)M(x - M^{-1}b) - b^T M^{-1} b$$

This leads to $R(w) - R(\hat{w}) = (w - \hat{w})^T A^T A(w - \hat{w})$

Describe geometrically the contour lines of $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(w) = R(w) - R(\hat{w})$