二次型梯度

benqbo 于 2018-11-14 16:01:28 发布

二次型梯度

最近看课本，开篇就是对向量求梯度，看来不弄清楚，第一页都跳不过，先来个定义，对向量 x 求梯度：

$$\nabla \mathrm{f}(\vec{x}) = \begin{pmatrix} \dfrac{df}{dx_1} \\ \dfrac{df}{dx_i} \\ \dfrac{df}{dx_n} \end{pmatrix}$$

1、向量 x 二范数平方的梯度：

$$\nabla(\|\mathrm{x}\|_2^2) = \nabla(x^{\mathrm{T}}x) = \begin{pmatrix} 2x_1 \\ 2x_{\mathrm{i}} \\ 2x_{\mathrm{n}} \end{pmatrix} = 2x$$

2、向量 x 带系数 b 的梯度：

$$\nabla(x^T b) = \nabla\left[(x_1, x_i, x_n)\begin{pmatrix} b_1 \\ b_i \\ b_n \end{pmatrix}\right] = \begin{pmatrix} b_1 \\ b_i \\ b_n \end{pmatrix} = \mathrm{b}$$

3、二次型梯度(A 是矩阵)：

$$\nabla(x^T \mathrm{A} x) = \nabla\left[(x_1, x_i, x_n)\begin{pmatrix} a_{11} & \dots & a_{1i} & a_{1n} \\ \dots & \dots & .. & .. \\ a_{i1} & .. & a_{ii} & a_{in} \\ a_{n1} & .. & a_{ni} & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_i \\ x_n \end{pmatrix}\right] = \nabla\left[\begin{pmatrix} x_1 a_{11} & x_1 a_{1i} & x_1 a_{1n} \\ + & + & + \\ x_i a_{i1} \dots & x_i a_{ii}, & x_i a_{in} \\ + & + & + \\ x_n a_{n1} & x_n a_{ni} & x_n a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_i \\ x_n \end{pmatrix}\right] = \nabla\left[x_1\begin{pmatrix} x_1 a_{11} \\ + \\ x_i a_{i1} \\ + \\ x_n a_{n1} \end{pmatrix} + \dots x_i\begin{pmatrix} x_1 a_{1i} \\ + \\ x_i a_{ii} \\ + \\ x_n a_{ni} \end{pmatrix} + x_n\begin{pmatrix} x_1 a_{1n} \\ + \\ x_i a_{in} \\ + \\ x_n a_{nn} \end{pmatrix}\right]$$

仔细观察，

$$x_1\begin{pmatrix} x_1 a_{11} \\ + \\ x_i a_{i1} \end{pmatrix} + \dots x_i\begin{pmatrix} x_1 a_{1i} \\ + \\ x_i a_{ii} \end{pmatrix} + x_n\begin{pmatrix} x_1 a_{1n} \\ + \\ x_i a_{in} \end{pmatrix}$$

$$\begin{pmatrix} \vdots \\ + \\ x_n a_{n1} \end{pmatrix} \quad \begin{pmatrix} \vdots \\ + \\ x_n a_{ni} \end{pmatrix} \quad \begin{pmatrix} \vdots \\ + \\ x_n a_{nn} \end{pmatrix}$$

其中含有 Xi 的项有：

$$x_i \begin{pmatrix} x_1 a_{1i} \\ + \\ x_i a_{ii} \\ + \\ x_n a_{ni} \end{pmatrix} = x_i (x_1 a_{1i} + \ldots x_i a_{ii} + x_n a_{ni})$$

$$x_1 \begin{pmatrix} \\ x_i a_{i1} \\ \\ \end{pmatrix} + \ldots x_i \begin{pmatrix} \\ x_i a_{ii} \\ \\ \end{pmatrix} + x_n \begin{pmatrix} \\ x_i a_{in} \\ \\ \end{pmatrix} = x_i (x_1 a_{i1} + \ldots x_i a_{ii} + x_n a_{in})$$

两项相加，注意 xi(xiaii) 重复加了一次，含有 Xi 的项：

$$x_i (x_1 a_{i1} + \ldots x_i a_{ii} + x_n a_{in}) + x_i (x_1 a_{1i} + \ldots x_i a_{ii} + x_n a_{ni}) - x_i x_i a_{ii}$$

对 xi 求导

$$\frac{\mathrm{d}\left(x^T \mathrm{A} x\right)}{\mathrm{d}x_i} = \frac{\mathrm{d}\left[x_i (x_1 a_{i1} + \ldots x_i a_{ii} + x_n a_{in}) + x_i (x_1 a_{1i} + \ldots x_i a_{ii} + x_n a_{ni}) - x_i x_i a_{ii}\right]}{\mathrm{d}x_i} = (x_1 a_{i1} + \ldots 2x_i a_{ii} + x_n a_{in}) + (x_1 a_{1i} + \ldots 2x_i a_{ii} + x_n a_{ni}) - 2x_i a_{ii}$$

$$= (x_1 a_{i1} + \ldots x_i a_{ii} + x_n a_{in}) + (x_1 a_{1i} + \ldots x_i a_{ii} + x_n a_{ni}) = (x_1 \ldots x_i \ldots x_n) \left[\begin{pmatrix} a_{i1} \\ a_{ii} \\ a_{in} \end{pmatrix} + \begin{pmatrix} a_{1i} \\ a_{ii} \\ a_{ni} \end{pmatrix} \right] = \left[(a_{i1} \ldots a_{ii} \ldots a_{in}) + (a_{1i} \ldots a_{ii} \ldots a_{ni}) \right] (x_1 \ldots x_i \ldots x_n)^T$$

得到

$$\nabla\left(x^T \mathrm{A} x\right) = \begin{pmatrix} \dfrac{\mathrm{d}\left(x^T \mathrm{A} x\right)}{\mathrm{d}x_1} \\ \dfrac{\mathrm{d}\left(x^T \mathrm{A} x\right)}{\mathrm{d}x_i} \\ \mathrm{d}\left(x^T \mathrm{A} x\right) \end{pmatrix} = \begin{pmatrix} (a_{11} \ldots a_{1i} .. a_{1n}) + (a_{11} \ldots a_{i1} .. a_{n1}) \\ (a_{i1} \ldots a_{ii} \ldots a_{in}) + (a_{1i} \ldots a_{ii} \ldots a_{ni}) \\ (a_{n1} \ldots a_{ni} .. a_{nn}) + (a_{1n} \ldots a_{in} .. a_{nn}) \end{pmatrix} (x_1 \ldots x_i \ldots x_n)^T = \left[\begin{pmatrix} a_{11} \ldots a_{1i} .. a_{1n} \\ a_{i1} \ldots a_{ii} \ldots a_{in} \\ a_{n1} \ldots a_{ni} .. a_{nn} \end{pmatrix} + \begin{pmatrix} a_{11} \ldots a_{1i} .. a_{1n} \\ a_{i1} \ldots a_{ii} \ldots a_{in} \\ a_{n1} \ldots a_{ni} .. a_{nn} \end{pmatrix}^T \right] (x_1 \ldots x_i \ldots x_n)^T$$

$$\left(\quad dx_n \quad \right)$$

即

$$\nabla\left(x^T \mathrm{A} x\right)=\left[\mathrm{A}+\mathrm{A}^T\right]\left(x_1 \ldots x_i \ldots x_n\right)^T=(\mathrm{A}+\mathrm{A}^T) x$$

如果 $\mathrm{A}=\mathrm{B}^{\mathrm{T}}\mathrm{B}$，有 $\mathrm{A}=\mathrm{B}^{\mathrm{T}}\mathrm{B}=(\mathrm{B}^{\mathrm{T}}\mathrm{B})^{\mathrm{T}}=\mathrm{A}^{\mathrm{T}}$，则

$$\nabla\left(x^T \mathrm{A} x\right)=2 \mathrm{A} x$$

4 例题

例 1 设 $\mathrm{A} \in \mathrm{R}^{\mathrm{n} \times \mathrm{n}}$，$\mathrm{b} \in \mathrm{R}^{\mathrm{n}}$ 且 $\Psi(x)=\frac{1}{2} x^T A x-x^T b$.证明 ψ 的梯度是 $\nabla \Psi(x)=\frac{1}{2}(\mathrm{A}^{\mathrm{T}}+\mathrm{A}) \mathrm{x}-\mathrm{b}$

证：由于前面已证明

$$\nabla\left(x^T \mathrm{A} x\right)=(\mathrm{A}+\mathrm{A}^T) x$$

$$\nabla\left(x^T \mathrm{b}\right)=\mathrm{b}$$

所以

$$\nabla \Psi(x)=\frac{1}{2} \nabla\left(x^T A x\right)-\nabla\left(x^T b\right)=\frac{1}{2}(A+A^T) x-b$$

例 2 设 $\mathrm{A} \in \mathrm{R}^{\mathrm{m} \times \mathrm{n}}$，则存在一个 n 维 2 范数单位向量 z，使得 $\mathrm{A}^{\mathrm{T}}\mathrm{A} \mathrm{z}=\mu^2 \mathrm{z}$，其中 $\mu=\|\mathrm{A}\|_2$.

例2利用矩阵范数的定义配合极值时梯度为0可证明