线性代数中的梯度计算

by 邢超(xingnix@live.com)

2016,9,5

1 从定义出发

1.1 $\boldsymbol{Y = X^T A\, X, A^T = A, d\,i\,m(A) = (N, N)\; d\,i\,m(X) = (N, 1)}$

$$\begin{aligned}
Y + d\,Y &= (X + d\,X)^T A\,(X + d\,X)\\
d\,Y &= d\,X^T A\,X + X^T A\,d\,X + d\,X^T A\,d\,X\\
d\,Y &= d\,X^T A\,X + X^T A\,d\,X + O\,(d\,X)\\
d\,Y &= X^T A\,d\,X + X^T A\,d\,X\\
d\,Y &= 2\,X^T A\,d\,X\\
\frac{d\,Y}{d\,X} &= 2\,X^T A
\end{aligned}$$

1.2 $\boldsymbol{E = (Y - A\,X)^T (Y - A\,X), d\,i\,m(A) = (M, N)\; d\,i\,m(X) = (N, 1)}$

$$\begin{aligned}
E + d\,E &= (Y - A\,(X + d\,X))^T (Y - A\,(X + d\,X))\\
&= (Y - A\,X - A\,d\,X)^T (Y - A\,X - A\,d\,X)\\
d\,E &= (-A\,d\,X)^T (Y - A\,X) + (Y - A\,X)^T (-A\,d\,X) + (A\,d\,X)^T (A\,d\,X)\\
&= (-A\,d\,X)^T (Y - A\,X) + (Y - A\,X)^T (-A\,d\,X) + O\,(d\,X)\\
&= (Y - A\,X)^T (-A\,d\,X) + (Y - A\,X)^T (-A\,d\,X)\\
&= 2\,(A\,X - Y)^T A\,d\,X\\
\frac{d\,E}{d\,X} &= 2\,(A\,X - Y)^T A
\end{aligned}$$

2 各分量偏导数

2.1 $\boldsymbol{Y = X^T A\, X, A^T = A, d\,i\,m(A) = (N, N)\; d\,i\,m(X) = (N, 1)}$

$$\begin{aligned}
Y &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i\, a_{i,j}\, x_j\\
\frac{\partial Y}{\partial x_k} &= \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial\,(x_i\, a_{i,j}\, x_j)}{\partial x_k}\\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (\frac{\partial x_i}{\partial x_k} a_{i,j}\, x_j + x_i\, a_{i,j} \frac{\partial x_j}{\partial x_k})\\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial x_i}{\partial x_k} a_{i,j}\, x_j + \sum_{i=1}^{N} \sum_{j=1}^{N} x_i\, a_{i,j} \frac{\partial x_j}{\partial x_k}\\
&= \sum_{j=1}^{N} a_{k,j}\, x_j + \sum_{i=1}^{N} x_i\, a_{i,k}\\
&= \sum_{i=1}^{N} a_{k,i}\, x_i + \sum_{i=1}^{N} x_i\, a_{i,k}\\
&= \sum_{i=1}^{N} (a_{k,i} + a_{i,k})\, x_i
\end{aligned}$$

$$= 2\sum_{i=1}^{N} a_{k,i} x_i$$

2.2 $\boldsymbol{E = (Y - AX)^T (Y - AX), dim(A) = (M, N)\ dim(X) = (N, 1)}$

$$\begin{aligned}
E &= \sum_{i=1}^{M} (y_i - \sum_{j=1}^{N} a_{i,j} x_j)^2 \\
\frac{E}{x_k} &= \sum_{i=1}^{M} \frac{\partial (y_i - \sum_{j=1}^{N} a_{i,j} x_j)^2}{\partial x_k} \\
&= 2\sum_{i=1}^{M} (y_i - \sum_{j=1}^{N} a_{i,j} x_j) \frac{\partial (y_i - \sum_{j=1}^{N} a_{i,j} x_j)}{\partial x_k} \\
&= 2\sum_{i=1}^{M} (y_i - \sum_{j=1}^{N} a_{i,j} x_j) \frac{\partial (-\sum_{j=1}^{N} a_{i,j} x_j)}{\partial x_k} \\
&= 2\sum_{i=1}^{M} (\sum_{j=1}^{N} a_{i,j} x_j - y_i) \sum_{j=1}^{N} \frac{\partial (a_{i,j} x_j)}{\partial x_k} \\
&= 2\sum_{i=1}^{M} (\sum_{j=1}^{N} a_{i,j} x_j - y_i) \sum_{j=1}^{N} a_{i,j} \frac{\partial x_j}{\partial x_k} \\
&= 2\sum_{i=1}^{M} (\sum_{j=1}^{N} a_{i,j} x_j - y_i) a_{i,k}
\end{aligned}$$

3 向量求导

3.1 $\boldsymbol{Y = AX}$

$$\frac{dY}{dX} = A$$

3.2 $\boldsymbol{Y = \mathbf{AU} + \mathbf{BV}, \mathbf{dim}\,(U) = (N, 1), \mathbf{dim}\,(V) = (M, 1)}$

$$\begin{aligned}
\frac{\partial Y}{\partial U} &= A \\
\frac{\partial Y}{\partial V} &= B
\end{aligned}$$

3.3 $\boldsymbol{Y = AU + BV, U = CX, V = DX, dim(X) = (N, 1)}$

$$\begin{aligned}
\frac{dY}{dX}\} &= \frac{\partial Y}{\partial U}\frac{dU}{dX} + \frac{\partial Y}{\partial V}\frac{dV}{dX} \\
&= AC + BD
\end{aligned}$$

3.4 $\boldsymbol{Y = X^T A X, A^T = A, dim(X) = (N, 1)}$

设

$$\begin{aligned}
X^T A X &= U^T A V \\
U &= X \\
V &= X
\end{aligned}$$

得

$$
\begin{aligned}
\frac{dY}{dX} &= \frac{\partial (U^T A V)}{\partial U}\frac{dU}{dX} + \frac{\partial (U^T A V)}{\partial V}\frac{dV}{dX} \\
&= \frac{\partial (V^T A^T U)}{\partial U}\frac{dU}{dX} + U^T A\frac{dV}{dX} \\
&= \frac{\partial (V^T A U)}{\partial U}\frac{dU}{dX} + U^T A\frac{dV}{dX} \\
&= V^T A + X^T A \\
&= X^T A + X^T A \\
&= 2X^T A
\end{aligned}
$$

3.5 $\boldsymbol{E = (Y - AX)^T (Y - AX), dim(A) = (M, N)\ dim(X) = (N, 1)}$

$$
\begin{aligned}
\frac{dY}{dX} &= \left(\frac{d(Y - AX)}{dX}\right)^T (Y - AX) + (Y - AX)^T \frac{d(Y - AX)}{dX} \\
&= (-A)^T (Y - AX) + (Y - AX)^T (-A) \\
&= (Y - AX)^T (-A) + (Y - AX)^T (-A) \\
&= 2(AX - Y)^T A
\end{aligned}
$$