线性代数中的梯度计算

邢超 (xingnix@live.com)

September 5, 2016

1 从定义出发

1.1 $Y = X^T AX, A^T = A, dim(A) = (N, N) dim(X) = (N, 1)$

$$
\begin{aligned}
Y + dY &= (X + dX)^T A(X + dX) \\
dY &= dX^T AX + X^T AdX + dX^T AdX \\
dY &= dX^T AX + X^T AdX + O(dX) \\
dY &= X^T AdX + X^T AdX \\
dY &= 2X^T AdX \\
\frac{dY}{dX} &= 2X^T A
\end{aligned}
$$

1.2 $E = (Y - AX)^T (Y - AX), dim(A) = (M, N) dim(X) = (N, 1)$

$$
\begin{aligned}
E + dE &= (Y - A(X + dX))^T (Y - A(X + dX)) \\
&= (Y - AX - AdX)^T (Y - AX - AdX) \\
dE &= (-AdX)^T (Y - AX) + (Y - AX)^T (-AdX) + (AdX)^T (AdX) \\
&= (-AdX)^T (Y - AX) + (Y - AX)^T (-AdX) + O(dX) \\
&= (Y - AX)^T (-AdX) + (Y - AX)^T (-AdX) \\
&= 2(AX - Y)^T AdX \\
\frac{dE}{dX} &= 2(AX - Y)^T A
\end{aligned}
$$

2 各分量偏导数

2.1 $Y = X^T AX, A^T = A, dim(A) = (N, N) dim(X) = (N, 1)$

$$
\begin{aligned}
Y &= \sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{i,j} x_j \\
\frac{\partial Y}{\partial x_k} &= \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial (x_i a_{i,j} x_j)}{\partial x_k} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} (\frac{\partial x_i}{\partial x_k} a_{i,j} x_j + x_i a_{i,j} \frac{\partial x_j}{\partial x_k}) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial x_i}{\partial x_k} a_{i,j} x_j + \sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{i,j} \frac{\partial x_j}{\partial x_k} \\
&= \sum_{j=1}^{N} a_{k,j} x_j + \sum_{i=1}^{N} x_i a_{i,k} \\
&= \sum_{i=1}^{N} a_{k,i} x_i + \sum_{i=1}^{N} x_i a_{i,k} \\
&= \sum_{i=1}^{N} (a_{k,i} + a_{i,k}) x_i \\
&= 2\sum_{i=1}^{N} a_{k,i} x_i
\end{aligned}
$$

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

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

3 向量求导

3.1 $Y = AX$

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

3.2 $Y = AU + BV, dim(U) = (N, 1), dim(V) = (M, 1)$

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

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

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

3.4 $Y = X^T AX, A^T = A, dim(X) = (N, 1)$

设

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

得

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

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

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