CSC321 Tutorial 1 partial derivative examples

Yue Li

Wed 11-12 Jan 15
Fri 10-11, Jan 17

Tutorial page:
http://www.cs.utoronto.ca/~yueli/CSC321_UTM_2014.html

Example 1

- $f = x^2 + b; \quad \frac{\partial f}{\partial x} = 2x$
- $f = x^k + b; \quad \frac{\partial f}{\partial x} = kx^{k-1}$
- $f = \log x; \quad \frac{\partial f}{\partial x} = \frac{1}{x}$
- $f = e^x; \quad \frac{\partial f}{\partial x} = e^x$

Example 2

Given response variable y and input variable x, we want to find w that minimizes E:

$y = wx$

$E = (y - wx)^2$ (what's the value of w to minimize E just by eye balling it?)

$$\frac{\partial E}{\partial w} = \frac{\partial (y - wx)^2}{\partial (y - wx)} \frac{\partial (y - wx)}{\partial w}$$
$$= 2(y - wx)(-x)$$

Let $\frac{\partial E}{\partial w} = 0$ and solve for w gives: $w = \frac{y}{x}$.

Example 3

$$
\begin{aligned}
E &= \sum_i (y_i - w_i x_i)^2 \\
\frac{\partial E}{\partial w_i} &= \frac{\partial E}{\partial w_i} \sum_{j \neq i} (y_j - w_j x_j)^2 + \frac{\partial E}{\partial w_i} (y_i - w_i x_i)^2 \\
\frac{\partial E}{\partial w_i} &= 0 + 2(y_i - w_i x_i)(-x_i) \\
\frac{\partial E}{\partial w_i} &= 2(y_i - w_i x_i)(-x_i)
\end{aligned}
$$

Setting $\frac{\partial E}{\partial w_i} = 0$ and solve for w_i gives: $w_i = \frac{y_i}{x_i}$.

Notably, $\frac{\partial E}{\partial w_i}$ (the partial derivative of E with respect to w_i) makes other terms not involving w_i become zero.

Example 4: meal plan pricing example

Suppose we only know the total price of the meal y_n and the portions of each dish 1,2,3 as x_i, $i \in \{1, 2, 3\}$. We want to find out the specific prices of each dish w_i. Here $n \in \{1 \dots N\}$ denotes each meal from day 1 to day N.

$$
\begin{aligned}
\hat{y}_n &= w_1 x_{1,n} + w_2 x_{2,n} + w_3 x_{3,n} \\
E &= \sum_{n=1}^{N} (y_n - w_1 x_{1,n} - w_2 x_{2,n} - w_3 x_{3,n})^2 \\
\frac{\partial E}{\partial w_i} &= \sum_n 2(y_n - w_1 x_{1,n} - w_2 x_{2,n} - w_3 x_{3,n})(-x_{i,n}) \\
&= \sum_n 2(y_n - \hat{y}_n)(-x_{i,n}) \equiv \Delta w_i
\end{aligned}
$$

Update $w_i^* = w_1 - \epsilon \Delta w_i$, where ϵ is the "learning rate" (e.g., $\epsilon = 10^{-5}$). With new w_i^*, we then repeat the above calculation to further update w_i^* and so forth. The iteration terminates when $E^{t-1} - E^t < tol$, tol is the threshold (e.g., $tol = 10^{-3}$), in which case the algorithm converges.