CALCULUS REVIEW

TO PREPARE
TO UNDERSTAND
NEURAL NETWORK
LEARNING

J. WUNDERLICH, Ph.D.

IF $f(x) = 2x \quad f'(x) = 2$

IF $f(x,y) = yx \quad \frac{\partial f}{\partial x} = y$

IF $f(x) = \frac{1}{2}(1-x)^2 \quad f'(x) = \frac{1}{2}\left(2(1-x)(-1)\right) = -(1-x)$

IF $f(x,y) = \frac{1}{2}(y-x)^2 \quad \frac{\partial f}{\partial x} = \frac{1}{2}\left(2(y-x)(-1)\right) = -(y-x)$

F $f(x) = 2 \quad f'(x) = 0$

F $f(x) = e^{-2x} \quad f'(x) = (e^{-2x})(-2) = -2e^{-2x}$

$$f(x) = \frac{x^2}{1-e^x} \quad f'(x) = ?$$

USE QUOTIENT RULE

$$f(x) = \frac{u(x)}{v(x)}$$

AND

$$f'(x) = \frac{v(x)u'(x) - u(x)v'(x)}{(v(x))^2}$$

FOR US

$$f(x) = \frac{u(x)}{v(x)} = \frac{x^2}{1-e^x}$$

$u(x) = x^2 \quad u'(x) = 2x$

$v(x) = 1-e^x \quad v'(x) = 0 - e^x$

$$f'(x) = \frac{(1-e^x)(2x) - x^2(-e^x)}{(1-e^x)^2}$$

$$= \frac{2x - 2xe^x + x^2e^x}{(1-e^x)^2}$$

Chain Rule

➤ *(Chain rule)* *If f and g are functions such that g is differentiable at x and f is differentiable at $u = g(x)$ and F is a function such that* $F(x) = f(g(x))$, *then*

$$\frac{d}{dx}F(x) = \frac{d}{du}f(u) \cdot \frac{d}{dx}g(x).$$

A simpler, but less accurate statement of this theorem is: If $y = f(u)$ and $u = g(x)$, then

$$\boxed{\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.}$$

Example 2

If $y = u^2$ and $u = x^2 - 4x + 3$, find dy/dx.

By the chain rule,
$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= 2u(2x - 4) \\ &= 4(x^2 - 4x + 3)(x - 2).\end{aligned}$$

Example 3

Find y' for $y = \dfrac{1}{(x^2 + 3x - 5)^3}$.

Let us make the substitution $u = x^2 + 3x - 5$. Then $y = u^{-3}$; and, by the chain rule,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= -3u^{-4}(2x + 3) \\ &= \frac{-3(2x + 3)}{(x^2 + 3x - 5)^4}.\end{aligned}$$

Example 4

Differentiate $y = \sqrt{x}$.

$$y = \sqrt{x} = x^{1/2}, \qquad y' = \frac{1}{2} x^{-1/2} = \frac{1}{2\sqrt{x}}.$$

Example 5

Differentiate $y = \dfrac{x+2}{\sqrt{x}}$.

There are several possible methods. Two are illustrated.

$$y = \frac{x+2}{x^{1/2}},$$

$$y' = \frac{x^{1/2} \cdot 1 - (x+2)\frac{1}{2}x^{-1/2}}{x}$$

$$= \frac{\sqrt{x} - \dfrac{x+2}{2\sqrt{x}}}{x}$$

$$= \frac{2x - (x+2)}{2x^{3/2}}$$

$$= \frac{x-2}{2x^{3/2}}.$$

$$y = x^{1/2} + 2x^{-1/2},$$

$$y' = \frac{1}{2}x^{-1/2} - x^{-3/2}$$

$$= \frac{1}{2x^{1/2}} - \frac{1}{x^{3/2}}$$

$$= \frac{x-2}{2x^{3/2}}.$$

In the second method of Example 5 we have avoided the relatively complicated quotient formula by carrying out the division and using negative exponents. While this method is not universally recommended, it sometimes simplifies a problem considerably.

Example 6

Differentiate $y = \dfrac{1}{x}$.

QUOTIENT RULE

$$y' = \frac{x \cdot 0 - 1 \cdot 1}{x^2} = \frac{-1}{x^2}.$$

However, by writing the original problem as $y = x^{-1}$, we get

SIMPLER: $$y' = -x^{-2} = -\frac{1}{x^2}.$$

The use of negative exponents makes the problem simple enough to do in your head. This method can be used to advantage when the denominator is very simple or when the numerator is a constant.

Recall that in Section 4.4 we considered the chain rule for differentiating a function of a function. For $y = f(u)$ and $u = g(x)$, we had $y = f(g(x))$ and

$$\boxed{\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.}$$

This proved to be very useful, especially when deriving formulas for the derivatives of transcendental functions. In this section we extend the chain rule to functions of several variables.

Theorem 21.2

Suppose $z = f(x, y)$, and $x(t)$ and $y(t)$ are differentiable functions in an open interval containing t. If $\partial z/\partial x$ and $\partial z/\partial y$ are continuous in a neighborhood of $(x(t), y(t))$, then

$$\boxed{\frac{dz}{dt} = \frac{\partial z}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial z}{\partial y} \cdot \frac{dy}{dt}.}$$

Example 1

If $z = x^2 + y^2$, $x = \sin t$ and $y = e^t$, find dz/dt.

$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\ &= 2x \cdot \cos t + 2y \cdot e^t \\ &= 2(x \cos t + ye^t). \end{aligned}$$

By substituting $x = \sin t$ and $y = e^t$, we can express the result entirely in terms of t.

$$\frac{dz}{dt} = 2(\sin t \cos t + e^{2t}).$$

Of course, this derivative could also be found by substituting first and then differentiating.

$$\begin{aligned} z &= x^2 + y^2 \\ &= \sin^2 t + e^{2t}; \\ \frac{dz}{dt} &= 2 \sin t \cos t + 2e^{2t}. \end{aligned}$$

Theorem 21.2 is especially useful when we do not know what all of the functions are. It can be extended in many ways. Perhaps the most obvious is the case in which z is a function of three or more variables, each of which is a function of t.

Theorem 21.3

If $z = f(x_1, x_2, \ldots, x_n)$ *and* $x_i = g_i(t)$, $i = 1, 2, \ldots, n$, *and if all partial derivatives of* z *are continuous and* dx_i/dt *exist,* $i = 1, 2, \ldots, n$, *then*

$$\frac{dz}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}.$$

A special case of Theorem 21.2 follows.

Theorem 21.4

If $z = f(x, y)$ *and* $y = g(x)$ *and if* $\partial z/\partial x$ *and* $\partial z/\partial y$ *are continuous and* dz/dx *exists, then*

$$\boxed{\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y}\cdot\frac{dy}{dx}.}$$

Example 2

If $z = x^2 + xy + y^2$ and $y = \sin x$, find dz/dx.

$$\frac{dz}{dx} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y}\cdot\frac{dy}{dx} = (2x + y) + (x + 2y)\cos x.$$

Again, this expression can be put entirely in terms of x, or it could be found by substituting $y = \sin x$ first and differentiating in the ordinary way.

Note the distinction between $\partial z/\partial x$ and dz/dx. $\partial z/\partial x$ is determined by the original function $z = f(x, y)$, where x and y are assumed to be two independent variables. The fact that $y = g(x)$ puts an additional restriction upon x and y does not enter into consideration when one is finding $\partial z/\partial x$ or $\partial z/\partial y$. This restriction *is* taken into account when finding dz/dx.

Just as Theorem 21.2, in which z is a function of two variables, can be extended to give Theorem 21.3, in which z is a function of n variables, it can also be extended from the case in which x and y are functions of a single variable t to the case in which x and y are functions of m variables. Let us consider one special case here.

If $z = f(x, y)$, $x = F(u, v)$, *and* $y = G(u, v)$ *and if* $\partial z/\partial x$ *and* $\partial z/\partial y$ *are continuous and* $\partial x/\partial u$, $\partial x/\partial v$, $\partial y/\partial u$, *and* $\partial y/\partial v$ *exist, then*

$$\boxed{\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial u}}$$

and

$$\boxed{\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial v}.}$$

The proof of this theorem is analogous to the proof of Theorem 21.2.

Example 3

Given $z = x^2 - y^3$, $x = u + v$, and $y = u - v$, find $\partial z/\partial u$ and $\partial z/\partial v$.

$$\begin{aligned}\frac{\partial z}{\partial u} &= \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial u}\\ &= 2x\cdot 1 - 3y^2\cdot 1\\ &= 2x - 3y^2.\end{aligned}$$

$$\begin{aligned}\frac{\partial z}{\partial v} &= \frac{\partial z}{\partial x}\cdot\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial v}\\ &= 2x\cdot 1 - 3y^2(-1)\\ &= 2x + 3y^2.\end{aligned}$$

A question that arises is, "How do we know when to use a partial derivative and when to use an ordinary derivative?" It is simply a matter of noting whether the function in question is a function of one variable or more than one. For instance, in Theorem 21.5, z is a function of the two variables x and y. Thus we want the partial derivatives

$$\frac{\partial z}{\partial x} \text{ and } \frac{\partial z}{\partial y}.$$

Again, both x and y are functions of the two variables u and v. We again want partial derivatives

$$\frac{\partial x}{\partial u},\ \frac{\partial x}{\partial v},\ \frac{\partial y}{\partial u},\ \frac{\partial y}{\partial v}.$$

Finally, if the x and y in $z = f(x, y)$ are replaced by $F(u, v)$ and $G(u, v)$, we have $z = f(F(u, v), G(u, v))$, which is still a function of two variables. Thus, we still want partial derivatives,

$$\frac{\partial z}{\partial u} \text{ and } \frac{\partial z}{\partial v}.$$

The chain rule is useful in rate-of-change problems.

$\boxed{\textit{Example 4}}$

Suppose the radius of a right circular cylinder is increasing at the rate of 2 in./min and the height is decreasing at 4 in./min. At what rate is the volume changing at the moment when the radius is 4 in. and the height is 10 in.?

$V = \pi r^2 h$. Since both r and h are functions of the time t we use the chain rule to differentiate.

$$\frac{dV}{dt} = \frac{\partial V}{\partial r}\frac{dr}{dt} + \frac{\partial V}{\partial h}\frac{dh}{dt}$$

$$= 2\pi r h \frac{dr}{dt} + \pi r^2 \frac{dh}{dt}.$$

But $dr/dt = 2$, $dh/dt = -4$, $r = 4$, and $h = 10$. Therefore

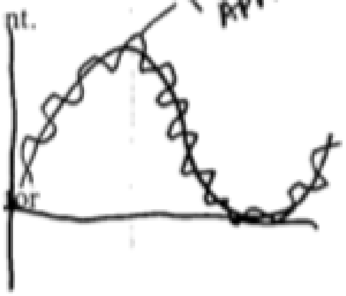
$$\frac{dV}{dt} = 2\pi(4)(10)(2) + \pi(4)^2(-4)$$

$$= 96\pi \text{ in.}^3/\text{min}.$$

19.5 Method of Least Squares

nt.

or



In **curve fitting** we are given n points (pairs of numbers)

$$(x_1, y_1), \cdots, (x_n, y_n)$$

FOR

and we want to determine a function $f(x)$ such that $f(x_j) \approx y_j$, $j = 1, \cdots, n$. The type of function (for example, polynomials, exponential functions, sine and cosine functions) may be suggested by the nature of the problem (the underlying physical law, for instance), and in many cases a polynomial of a certain degree will be AT appropriate. APPROXIMATED

If we require strict equality $f(x_1) = y_1, \cdots, f(x_n) = y_n$ and use polynomials of sufficiently high degree, we may apply one of the methods discussed in Sec. 18.3 in connection with interpolation. However, in certain situations this would not be the appropriate solution of the actual problem. For instance, to the four points

$$(1) \qquad (-1.0,\ 1.000), \qquad (-0.1,\ 1.099), \qquad (0.2,\ 0.808), \qquad (1.0,\ 1.000)$$

there corresponds the Lagrange polynomial $f(x) = x^3 - x + 1$ (Fig. 432), but if we graph the points, we see that they lie nearly on a straight line. Hence if these values are obtained in an experiment and thus involve an experimental error, and if the nature of the experiment suggests a linear relation, we better fit a straight line through the points (Fig. 432). Such a line may be useful for predicting values to be expected for other values of x. In simple cases a straight line may be fitted by eye, but if the points are scattered, this becomes unreliable and we better use a mathematical principle. A widely used procedure of this type is the **method of least squares** by Gauss. In the present situation it may be formulated as follows.

CURVE FIT A STRAIGHT LINE!

Method of least squares. *The straight line*

$$y = b + mx$$

should be fitted through the given points $(x_1, y_1), \cdots, (x_n, y_n)$ *so that the*

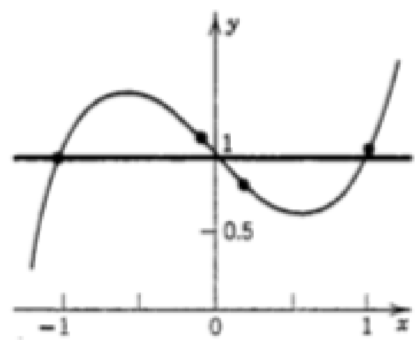


Fig. 432. The approximate fitting of a straight line

sum of the squares of the distances of those points from the straight line is minimum, where the distance is measured in the vertical direction (the y-direction).

The point on the line with abscissa x_j has the ordinate $b + mx_j$. Hence its distance from (x_j, y_j) is $|y_j - b - mx_j|$ (cf. Fig. 433) and that sum of squares is

$$\boxed{q = \sum_{j=1}^{n} (y_j - b - mx_j)^2.}$$

q depends on b and m. A necessary condition for q to be minimum is

$$\begin{aligned}
(2) \qquad \frac{\Delta q}{\Delta b} \propto \quad & \frac{\partial q}{\partial b} = -2 \sum_{j=1}^{n} (y_j - b - mx_j) = 0 \\
\frac{\Delta q}{\Delta m} \propto \quad & \frac{\partial q}{\partial m} = -2 \sum_{j=1}^{n} x_j(y_j - b - mx_j) = 0
\end{aligned}$$

(where we sum over j from 1 to n). Writing each sum as three sums and taking one of them to the right, we obtain the result

$$(3) \qquad \boxed{\begin{aligned} bn \quad + m\sum x_j &= \sum y_j \\ b\sum x_j + b\sum x_j^2 &= \sum x_j y_j. \end{aligned}}$$

These equations are called the **normal equations** of our problem.

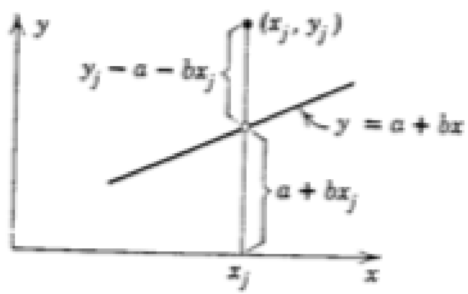


Fig. 433. Vertical distance of a point (x_j, y_j) from a straight line $y = a + bx$

EXAMPLE 1. Straight line

Using the method of least squares, fit a straight line to the four points given in formula (1).

Solution. We obtain

$$n = 4, \quad \sum x_j = 0.1, \quad \sum x_j^2 = 2.05, \quad \sum y_j = 3.907, \quad \sum x_j y_j = 0.0517.$$

Hence the normal equations are

$$4a + 0.10b = 3.9070$$

$$0.1a + 2.05b = 0.0517$$

The solution is $a = 0.9773$, $b = -0.0224$, and we obtain the straight line (Fig. 432)

$$y = 0.9773 - 0.0224x.$$