Taylor series and linear approximation

Ten problems on tangent lines and Taylor series — the sign of a linear approximation's error, the standard Maclaurin series, series by substitution, a Taylor polynomial about 4, a Lagrange remainder bound, a radius of convergence, and the softplus, log-sigmoid and quadratic-loss expansions that machine learning uses — with worked solutions and the mistakes that drop the n! or use a series outside its radius.

BEFORE YOU START

A Taylor polynomial replaces a function near a point by the polynomial with the same value and first few derivatives there: the tangent line is the degree-1 case, and the quadratic model behind Newton's method is the degree-2 case. The work goes wrong in a few places that recur: a coefficient without its $n!$, powers of x where powers of $x-a$ belong, a lost alternating sign, a remainder bound that takes the derivative's maximum in the wrong place, and a series used outside its radius. These ten problems run from a tangent line to the standard series, error bounds, and the expansions of softplus, the log-sigmoid and a loss.

- $T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$ is the degree-n Taylor polynomial of f about a, where $f^{(k)}$ is the k-th derivative and $0! = 1$. About $a = 0$ it is a Maclaurin polynomial; T_1 is the tangent line.
- Lagrange remainder: $f(x) - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ for some unknown c between a and x, so a bound uses the largest $|f^{(n+1)}|$ between them.
- A power series $\sum c_n(x-a)^n$ converges for $|x-a| < R$, its radius of convergence, and diverges for $|x-a| > R$. Inside the radius it can be integrated term by term, and it is the only power series of its sum about a. Ratio test: if $|b_{n+1}/b_n| \to \ell$, then $\sum b_n$ converges for $\ell < 1$ and diverges for $\ell > 1$.
- $\sigma(z) = 1/(1+e^{-z})$ and $\operatorname{softplus}(z) = \ln(1+e^{z})$, with $\operatorname{softplus}' = \sigma$ and $\sigma' = \sigma(1-\sigma)$ from the previous pages. $O(z^k)$ stands for terms of degree k and higher; angles are in radians.

Builds on: Differentiation rules: chain, product, quotient

Problems

1. Write the tangent line of $f(x) = \sqrt{x}$ at $a = 4$, use it to approximate $\sqrt{4.1}$, and decide from the concavity of f whether the approximation is too high or too low. `difficulty ·`

2. Find the Maclaurin series of e^x, $\sin x$ and $\cos x$ from their derivatives at 0, and write each with a general term. difficulty ·

3. Starting from the geometric series, find the Maclaurin series of $\dfrac{1}{1-x}$ and of $\ln(1+x)$, and say for which x they hold. difficulty ··

4. Use known series to write the Maclaurin series of e^{-x^2} through the x^6 term and of $x\cos x$ through the x^5 term. Then read off the sixth derivative of e^{-x^2} at $x=0$. difficulty ··

5. Find the degree-2 Taylor polynomial of $\sqrt{x}$ about $a=4$ and use it to improve the estimate of $\sqrt{4.1}$ from Problem 1. difficulty ··

6. Let T_n be the degree-n Maclaurin polynomial of $\ln(1+x)$. Use the Lagrange remainder to find the smallest n for which it guarantees $|\ln 1.5 - T_n(0.5)| < 10^{-3}$. difficulty ···

7. (a) Show that $|\sin\theta - \theta| \le |\theta|^3/6$, and bound the error of the small-angle approximation $\sin\theta \approx \theta$ for $|\theta| \le 0.1$. (b) Expand $\sqrt{1+x}$ about 0 to second order and estimate $\sqrt{1.02}$. difficulty ··

8. Find the radius of convergence of $\displaystyle\sum_{n\ge 1} \frac{x^n}{n\,3^n}$ and the function it sums inside that radius. difficulty ··

9. Expand $\operatorname{softplus}(z) = \ln(1+e^z)$ about $z=0$ through the z^2 term, and show that its z^3 term is 0. Then give the first-order expansion of $\ln\sigma(z)$ about $z=0$. difficulty ···

10. A scalar loss f is twice differentiable at w; let $g = f'(w)$ and $h = f''(w)$. Write the second-order Taylor model of $f(w+\delta)$ in the step δ, find the step δ^* that minimizes the model when $h>0$, and evaluate both for the logistic loss $L(w)=\ln(1+e^{-w})$ at $w=0$. (In many variables the same model reads $f(\mathbf{w}+\boldsymbol{\delta}) \approx f(\mathbf{w}) + \mathbf{g}^\top\boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^\top H\boldsymbol{\delta}$; this problem is its one-variable case.) difficulty ···

Answers

1. $L(x) = 2 + \frac{1}{4}(x-4)$, so $\sqrt{4.1} \approx 2.025$, an overestimate
2. $e^x = \sum_{n\ge 0} \dfrac{x^n}{n!}$, $\sin x = \sum_{n\ge 0} \dfrac{(-1)^n x^{2n+1}}{(2n+1)!}$, $\cos x = \sum_{n\ge 0} \dfrac{(-1)^n x^{2n}}{(2n)!}$
3. $\dfrac{1}{1-x} = \sum_{n\ge 0} x^n$ for $|x| < 1$; $\ln(1+x) = \sum_{n\ge 1} \dfrac{(-1)^{n+1} x^n}{n}$ for $-1 < x \le 1$
4. $e^{-x^2} = 1 - x^2 + \frac{x^4}{2} - \frac{x^6}{6} + \cdots$; $x\cos x = x - \frac{x^3}{2} + \frac{x^5}{24} - \cdots$; $\left.\frac{d^6}{dx^6} e^{-x^2}\right|_{x=0} = -120$
5. $T_2(x) = 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2$, and $T_2(4.1) = 2.02484375$
6. $n = 7$
7. $|\sin\theta - \theta| \le \frac{|\theta|^3}{6} \le 1.7 \times 10^{-4}$ for $|\theta| \le 0.1$; $\sqrt{1+x} \approx 1 + \frac{x}{2} - \frac{x^2}{8}$, so $\sqrt{1.02} \approx 1.00995$
8. $R = 3$, and $\sum_{n\ge 1} \dfrac{x^n}{n 3^n} = -\ln\!\left(1 - \frac{x}{3}\right)$ for $|x| < 3$
9. $\ln(1 + e^z) = \ln 2 + \frac{z}{2} + \frac{z^2}{8} + O(z^4)$, and $\ln\sigma(z) = -\ln 2 + \frac{z}{2} + O(z^2)$
10. $\delta^* = -\dfrac{f'(w)}{f''(w)}$; for L at $w = 0$, $L(\delta) \approx \ln 2 - \frac{\delta}{2} + \frac{\delta^2}{8}$ and $\delta^* = 2$

Worked solutions

Problem 1

> Write the tangent line of $f(x) = \sqrt{x}$ at $a = 4$, use it to approximate $\sqrt{4.1}$, and decide from the concavity of f whether the approximation is too high or too low.

$f(4) = 2$; $f'(x) = \dfrac{1}{2\sqrt{x}}$, so $f'(4) = \frac{1}{4}$.	The tangent line needs the value and the slope of f at $a = 4$; $\sqrt{x} = x^{1/2}$, and the power rule gives $f'(x) = \frac{1}{2}x^{-1/2}$.
$L(x) = f(4) + f'(4)(x - 4) = 2 + \frac{1}{4}(x - 4)$.	The tangent line is T_1, the degree-1 Taylor polynomial about 4: it matches f and f' there.
$L(4.1) = 2 + \frac{1}{4} \cdot 0.1 = 2.025$.	$x - 4 = 0.1$.
$f''(x) = -\frac{1}{4}x^{-3/2} < 0$ for $x > 0$.	Power rule on $\frac{1}{2}x^{-1/2}$; $x^{-3/2} > 0$, so the sign comes from $-\frac{1}{4}$ and f is concave.
$f(x) - L(x) = \frac{1}{2}f''(c)(x - 4)^2 < 0$ for some c between 4 and $x \neq 4$.	Lagrange remainder with $n = 1$: $f''(c) < 0$ by step 4 and $(x - 4)^2 > 0$, so a concave curve lies below its tangent line on both sides of the point of contact.
$L(x) = 2 + \frac{1}{4}(x - 4)$, so $\sqrt{4.1} \approx 2.025$, an overestimate	Step 5 at $x = 4.1$. Indeed $\sqrt{4.1} = 2.024846\ldots$, about 1.5×10^{-4} below the tangent line.

Problem 2

> Find the Maclaurin series of e^x, $\sin x$ and $\cos x$ from their derivatives at 0, and write each with a general term.

Every derivative of e^x is e^x, so $\frac{d^n}{dx^n}e^x\big\vert_{x=0} = e^0 = 1$ for every n.	$(e^x)' = e^x$, applied n times.
$e^x = 1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots$	Each coefficient is the derivative at 0, which is 1, divided by $n!$.

The derivatives of $\sin x$ cycle through $\cos x$, $-\sin x$, $-\cos x$, $\sin x$; at 0, starting from $\sin 0$, the values are $0, 1, 0, -1$, repeating.	$(\sin x)' = \cos x$ and $(\cos x)' = -\sin x$; $\sin 0 = 0$ and $\cos 0 = 1$.
$\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \cdots$	The even derivatives are 0, so only odd powers appear; the odd ones alternate $1, -1$, and each is divided by its factorial.
$\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \cdots$	$\cos x$ is the derivative of $\sin x$, so its values at 0 are the cycle of step 3 shifted by one, $1, 0, -1, 0$: only even powers, alternating in sign.
$e^x = \sum_{n\ge 0} \dfrac{x^n}{n!}$, $\sin x = \sum_{n\ge 0} \dfrac{(-1)^n x^{2n+1}}{(2n+1)!}$, $\cos x = \sum_{n\ge 0} \dfrac{(-1)^n x^{2n}}{(2n)!}$	Count the nonzero terms with n: the n-th term of sin has degree $2n+1$ and of cos degree $2n$, each with sign $(-1)^n$. All three equal their functions for every x: the Lagrange remainder is at most $e^{\lvert x\rvert}\lvert x\rvert^{n+1}/(n+1)!$ for e^x and $\lvert x\rvert^{n+1}/(n+1)!$ for sin and cos, and $(n+1)!$ outgrows $\lvert x\rvert^{n+1}$.

Problem 3

> Starting from the geometric series, find the Maclaurin series of $\dfrac{1}{1-x}$ and of $\ln(1+x)$, and say for which x they hold.

$\dfrac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n\ge 0} x^n$ for $\lvert x\rvert < 1$.	The geometric series with ratio $r = x$: the partial sum $(1 - x^{N+1})/(1-x)$ tends to $1/(1-x)$ exactly when $\lvert x\rvert < 1$.
$\dfrac{1}{1+t} = \sum_{n\ge 0}(-t)^n = \sum_{n\ge 0}(-1)^n t^n$ for $\lvert t\rvert < 1$.	Step 1 with $-t$ in place of x: $1 - (-t) = 1 + t$, $\lvert -t\rvert = \lvert t\rvert$, and $(-t)^n = (-1)^n t^n$ carries the alternating sign.

$\ln(1+x) = \displaystyle\int_0^x \frac{dt}{1+t} = \sum_{n\geq 0} (-1)^n \frac{x^{n+1}}{n+1}$ for $\lvert x\rvert < 1$.	$\ln(1+x)$ is the antiderivative of $1/(1+t)$ that is 0 at $x=0$; a power series can be integrated term by term inside its radius, and $\int_0^x t^n\,dt = x^{n+1}/(n+1)$.
$\dfrac{1}{1-x} = \sum_{n\geq 0} x^n$ **for** $\lvert x\rvert < 1$; $\ln(1+x) = \sum_{n\geq 1} \dfrac{(-1)^{n+1}x^n}{n}$ **for** $-1 < x \leq 1$	Rename $n+1$ as n in step 3, which turns the sign $(-1)^n$ into $(-1)^{n+1}$: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$. Term-by-term integration gives $\lvert x\rvert < 1$; the log series also converges at $x=1$, to $\ln 2$ (the alternating series test, and Abel's theorem says the sum is $\ln 2$), but not at $x=-1$, where it is minus the harmonic series and $\ln 0$ is undefined.

Another route: from the derivatives

$f(x) = \ln(1+x)$: $f'(x) = (1+x)^{-1}$, $f''(x) = -(1+x)^{-2}$, $f'''(x) = 2(1+x)^{-3}$.	Chain rule with inner derivative 1; each differentiation multiplies by the current exponent and lowers it by 1.
$f^{(n)}(x) = (-1)^{n+1}(n-1)!\,(1+x)^{-n}$ for $n \geq 1$.	Differentiating $(1+x)^{-n}$ brings down $-n$, which flips the sign and turns $(n-1)!$ into $n!$: the formula for n gives the formula for $n+1$.
$\dfrac{f^{(n)}(0)}{n!} = \dfrac{(-1)^{n+1}(n-1)!}{n!} = \dfrac{(-1)^{n+1}}{n}$, and $f(0) = \ln 1 = 0$.	The Taylor coefficient divides the derivative by $n!$, and $n! = n\cdot(n-1)!$.
$\ln(1+x) = \sum_{n\geq 1} \dfrac{(-1)^{n+1}x^n}{n}$, **for** $\lvert x\rvert < 1$	The coefficients agree with the first route. The ratio of consecutive terms is $\lvert x\rvert \cdot \frac{n}{n+1} \to \lvert x\rvert$, so the ratio test gives convergence for $\lvert x\rvert < 1$.

Problem 4

> Use known series to write the Maclaurin series of e^{-x^2} through the x^6 term and of $x\cos x$ through the x^5 term. Then read off the sixth derivative of e^{-x^2} at $x = 0$.

$e^{u} = 1 + u + \dfrac{u^2}{2} + \dfrac{u^3}{6} + \cdots$ with $u = -x^2$.	Problem 2's series holds for every u, so it holds for $u = -x^2$; no derivative of e^{-x^2} has to be computed.
$e^{-x^2} = 1 - x^2 + \dfrac{x^4}{2} - \dfrac{x^6}{6} + \cdots$	$(-x^2)^2 = x^4$ and $(-x^2)^3 = -x^6$.
$x\cos x = x\left(1 - \dfrac{x^2}{2} + \dfrac{x^4}{24} - \cdots\right) = x - \dfrac{x^3}{2} + \dfrac{x^5}{24} - \cdots$	Multiplying a convergent series by x multiplies each term by x; $2! = 2$ and $4! = 24$.
For $f(x) = e^{-x^2}$, $\dfrac{f^{(6)}(0)}{6!} = -\dfrac{1}{6}$.	A function has only one power series about 0, so the x^6 coefficient in step 2 is the Taylor coefficient $f^{(6)}(0)/6!$.
$e^{-x^2} = 1 - x^2 + \frac{x^4}{2} - \frac{x^6}{6} + \cdots$; $x\cos x = x - \frac{x^3}{2} + \frac{x^5}{24} - \cdots$; $\frac{d^6}{dx^6} e^{-x^2}\big\vert_{x=0} = -120$	$6! \cdot (-\frac{1}{6}) = -720/6 = -120$. Differentiating e^{-x^2} six times by the chain and product rules reaches the same number by a much longer road.

Problem 5

> Find the degree-2 Taylor polynomial of $\sqrt{x}$ about $a = 4$ and use it to improve the estimate of $\sqrt{4.1}$ from Problem 1.

$f(4) = 2$, $f'(4) = \frac{1}{4}$, and $f''(4) = -\frac{1}{4} \cdot 4^{-3/2} = -\frac{1}{32}$.	Problem 1, steps 1 and 4; $4^{3/2} = 8$.
$T_2(x) = f(4) + f'(4)(x-4) + \dfrac{f''(4)}{2!}(x-4)^2$.	About $a = 4$ the polynomial is in powers of $x - 4$, the distance from the point where the derivatives were taken, and the k-th derivative is divided by $k!$.
$\dfrac{f''(4)}{2!} = -\dfrac{1}{64}$.	$-\frac{1}{32}$ divided by 2.

$T_2(4.1) = 2 + 0.025 - \dfrac{0.01}{64} = 2.025 - 0.00015625.$	$x - 4 = 0.1$ and $(0.1)^2 = 0.01$. The new term corrects the tangent line downward, as Problem 1's concavity argument requires.
$\lvert\sqrt{4.1} - T_2(4.1)\rvert \le \dfrac{3/256}{3!}(0.1)^3 < 2 \times 10^{-6}.$	Lagrange remainder with $n = 2$: $f'''(x) = \frac{3}{8}x^{-5/2}$ is decreasing, so on $[4, 4.1]$ it is largest at $x = 4$, where it is $\frac{3}{8} \cdot \frac{1}{32} = \frac{3}{256}$; and $\frac{3}{256} \cdot \frac{0.001}{6} \approx 1.95 \times 10^{-6}$.
$T_2(x) = 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2$, **and** $T_2(4.1) = 2.02484375$	Step 4. By step 5 the error is below 2×10^{-6}; $\sqrt{4.1} = 2.0248457\ldots$, so it is about 1.9×10^{-6}, eighty times smaller than the tangent line's.

Problem 6

> Let T_n be the degree-n Maclaurin polynomial of $\ln(1+x)$. Use the Lagrange remainder to find the smallest n for which it guarantees $\lvert\ln 1.5 - T_n(0.5)\rvert < 10^{-3}$.

$f^{(n+1)}(c) = (-1)^n\, n!\,(1+c)^{-(n+1)}$ for $f(x) = \ln(1+x)$.	Problem 3, alt route, step 2, with $n+1$ in place of n.
$\ln 1.5 - T_n(0.5) = \dfrac{f^{(n+1)}(c)}{(n+1)!}(0.5)^{n+1}$ for some c in $(0, 0.5)$.	Lagrange remainder with $a = 0$ and $x = 0.5$; all that is known about c is that it lies between them.
$\lvert f^{(n+1)}(c)\rvert = \dfrac{n!}{(1+c)^{n+1}} \le n!$ for $0 < c < 0.5$.	The bound has to hold for whichever c the theorem supplies, so it uses the largest value on the interval; $(1+c)^{-(n+1)}$ is largest where $1+c$ is smallest, at $c = 0$.
$\lvert\ln 1.5 - T_n(0.5)\rvert \le \dfrac{n!}{(n+1)!}(0.5)^{n+1} = \dfrac{0.5^{n+1}}{n+1}.$	Steps 2 and 3, with $(n+1)! = (n+1) \cdot n!$.

$n=6$: $\dfrac{0.5^7}{7}=\dfrac{1}{896}\approx 1.1\times10^{-3}$; $n=7$: $\dfrac{0.5^8}{8}=\dfrac{1}{2048}\approx 4.9\times10^{-4}$.	The bound decreases as n grows, so the first n at which it drops below 10^{-3} is the smallest one it guarantees.
$n=7$	Step 5. The guarantee is conservative: the actual error of $T_6(0.5)$ is already 7.8×10^{-4}, but the bound cannot see that, and it is the bound that the question asks for.

Problem 7

(a) Show that $|\sin\theta-\theta|\le|\theta|^3/6$, and bound the error of the small-angle approximation $\sin\theta\approx\theta$ for $|\theta|\le 0.1$. (b) Expand $\sqrt{1+x}$ about 0 to second order and estimate $\sqrt{1.02}$.

For $\sin$ about 0, $T_2(\theta)=\theta$.	Problem 2: the θ^2 coefficient of $\sin$ is 0, so T_1 and T_2 coincide, and the sharper remainder after degree 2 applies to $\sin\theta\approx\theta$.
$\sin\theta-\theta=\dfrac{-\cos c}{3!}\theta^3$ for some c between 0 and θ.	Lagrange remainder with $n=2$; the third derivative of $\sin$ is $-\cos$.
$\lvert\sin\theta-\theta\rvert\le\dfrac{\lvert\theta\rvert^3}{6}\le\dfrac{0.001}{6}\approx1.67\times10^{-4}$ for $\lvert\theta\rvert\le0.1$.	$\lvert\cos c\rvert\le1$ for every c. Relative to θ the error is at most $\theta^2/6\le0.17\%$.
For $f(x)=(1+x)^{1/2}$: $f(0)=1$, $f'(0)=\frac12$, $f''(0)=-\frac14$.	$f'(x)=\frac12(1+x)^{-1/2}$ and $f''(x)=-\frac14(1+x)^{-3/2}$, both evaluated at $x=0$.
$\sqrt{1+x}\approx1+\dfrac{x}{2}-\dfrac{x^2}{8}$.	T_2 about 0; $f''(0)/2!=-\frac18$.
$\sqrt{1.02}\approx1+0.01-\dfrac{0.0004}{8}=1.00995$.	$x=0.02$ and $x^2=0.0004$.

$\lvert\sin\theta - \theta\rvert \le \frac{\lvert\theta\rvert^3}{6} \le 1.7 \times 10^{-4}$ **for** $\lvert\theta\rvert \le 0.1$; $\sqrt{1+x} \approx 1 + \frac{x}{2} - \frac{x^2}{8}$, **so** $\sqrt{1.02} \approx 1.00995$	Steps 3, 5 and 6. $\sqrt{1.02} = 1.0099505\ldots$; the error, 5×10^{-7}, is the size of the next term, $f'''(0)x^3/3! = x^3/16$ at $x = 0.02$.

Problem 8

> Find the radius of convergence of $\displaystyle\sum_{n\ge 1} \frac{x^n}{n\,3^n}$ and the function it sums inside that radius.

$b_n = \dfrac{x^n}{n3^n}$ and $\left\lvert\dfrac{b_{n+1}}{b_n}\right\rvert = \dfrac{\lvert x\rvert}{3}\cdot\dfrac{n}{n+1}.$	The ratio test needs the ratio of consecutive terms: $x^{n+1}/x^n = x$, $3^n/3^{n+1} = \frac{1}{3}$ and $n/(n+1)$ from the n in the denominator.
$\dfrac{\lvert x\rvert}{3}\cdot\dfrac{n}{n+1} \to \dfrac{\lvert x\rvert}{3}$ as $n \to \infty$.	$n/(n+1) = 1/(1+1/n) \to 1$.
$R = 3$.	By the ratio test the series converges when $\lvert x\rvert/3 < 1$ and diverges when $\lvert x\rvert/3 > 1$.
$\ln\!\left(1 - \frac{x}{3}\right) = \sum_{n\ge1} \dfrac{(-1)^{n+1}(-x/3)^n}{n} = -\sum_{n\ge1} \dfrac{x^n}{n3^n}$ for $\lvert x\rvert < 3$.	Problem 3's series with $-x/3$ in place of x, valid when $\lvert x/3\rvert < 1$; $(-1)^{n+1}(-1)^n = -1$ for every n.
$R = 3$, **and** $\sum_{n\ge1} \dfrac{x^n}{n3^n} = -\ln\!\left(1 - \frac{x}{3}\right)$ **for** $\lvert x\rvert < 3$	Steps 3 and 4, with step 4 negated. At $x = 3$ the series is the harmonic series and diverges; at $x = -3$ it is the alternating harmonic series and converges, to $-\ln 2$.

Problem 9

> Expand $\operatorname{softplus}(z) = \ln(1 + e^z)$ about $z = 0$ through the z^2 term, and show that its z^3 term is 0. Then give the first-order expansion of $\ln\sigma(z)$ about $z = 0$.

$\operatorname{softplus}(0) = \ln 2$.	$1 + e^0 = 2$.

$\mathrm{softplus}'(0) = \sigma(0) = \frac{1}{2}$ and $\mathrm{softplus}''(0) = \sigma(0)\big(1-\sigma(0)\big) = \frac{1}{4}$.	$\mathrm{softplus}' = \sigma$ and $\sigma' = \sigma(1-\sigma)$ (Differentiation rules, Problems 4 and 5); $\sigma(0) = 1/(1+1)$.
$\mathrm{softplus}(z) = \ln 2 + \dfrac{z}{2} + \dfrac{z^2}{8} + \cdots$	Taylor coefficients about 0: $\frac{1}{2}$ for z, and $\frac{1}{4}/2! = \frac{1}{8}$ for z^2.
$\mathrm{softplus}(z) - \mathrm{softplus}(-z) = \ln \dfrac{1+e^{z}}{1+e^{-z}} = \ln e^{z} = z.$	The difference of logs is the log of the quotient, and $1 + e^{z} = e^{z}(e^{-z}+1)$, so the quotient is e^{z}.
$\mathrm{softplus}(z) - \frac{z}{2}$ is an even function, so the z^3 coefficient of softplus is 0.	Step 4 rearranges to $\mathrm{softplus}(z) - \frac{z}{2} = \mathrm{softplus}(-z) - \frac{-z}{2}$. An even function's series has only even powers, so every odd power of softplus comes from $\frac{z}{2}$, which has no z^3 term.
$\ln \sigma(z) = \ln \dfrac{1}{1+e^{-z}} = -\,\mathrm{softplus}(-z).$	The log of a reciprocal is the negative of the log.
$-\,\mathrm{softplus}(-z) = -\ln 2 + \dfrac{z}{2} - \dfrac{z^2}{8} + \cdots$	Step 3 with $-z$ in place of z, then negated: $-\frac{-z}{2} = \frac{z}{2}$.
$\ln(1+e^{z}) = \ln 2 + \frac{z}{2} + \frac{z^2}{8} + O(z^4)$, **and** $\ln \sigma(z) = -\ln 2 + \frac{z}{2} + O(z^2)$	Steps 3, 5 and 7. At a logit of 0 the logistic loss $-\ln \sigma(z)$ is $\ln 2 \approx 0.693$, the loss of a classifier that outputs probability $\frac{1}{2}$, and it falls by about $\frac{1}{2}$ per unit increase of the logit.

Problem 10

> A scalar loss f is twice differentiable at w; let $g = f'(w)$ and $h = f''(w)$. Write the second-order Taylor model of $f(w+\delta)$ in the step δ, find the step δ^{*} that minimizes the model when $h > 0$, and evaluate both for the logistic loss $L(w) = \ln(1+e^{-w})$ at $w = 0$. (In many variables the same model reads $f(\mathbf{w}+\boldsymbol{\delta}) \approx f(\mathbf{w}) + \mathbf{g}^{\top}\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^{\top} H \boldsymbol{\delta}$; this problem is its one-variable case.)

$f(w+\delta) \approx q(\delta) = f(w) + g\delta + \frac{1}{2}h\delta^2.$	T_2 about w, evaluated at $w+\delta$: the powers of $x-w$ become powers of δ, and the second derivative is divided by 2!.

$q'(\delta) = g + h\delta = 0$ at $\delta = -g/h$.	With $h > 0$ the model is an upward-opening parabola in δ, so its one stationary point is its minimum.
$L'(w) = \dfrac{-e^{-w}}{1 + e^{-w}} = -\sigma(-w)$.	Chain rule: the derivative of $\ln$ at $1 + e^{-w}$ times $(1 + e^{-w})' = -e^{-w}$; dividing top and bottom by e^{-w} gives $1/(e^{w} + 1) = \sigma(-w)$.
$L''(w) = \sigma(-w)\big(1 - \sigma(-w)\big) = \sigma(w)\sigma(-w)$.	Chain rule on $-\sigma(-w)$: $\sigma'(-w) = \sigma(-w)(1 - \sigma(-w))$, and the inner derivative -1 cancels the leading minus; $1 - \sigma(-w) = \sigma(w)$.
$L(0) = \ln 2$, $g = L'(0) = -\frac{1}{2}$, $h = L''(0) = \frac{1}{4}$.	$\sigma(0) = \frac{1}{2}$.
$q(\delta) = \ln 2 - \dfrac{\delta}{2} + \dfrac{\delta^2}{8}$ and $\delta^* = -\dfrac{-1/2}{1/4} = 2$.	Steps 1, 2 and 5; $h = \frac{1}{4} > 0$, so the minimum exists.
$\delta^* = -\dfrac{f'(w)}{f''(w)}$; **for** L **at** $w = 0$, $L(\delta) \approx \ln 2 - \frac{\delta}{2} + \frac{\delta^2}{8}$ **and** $\delta^* = 2$	Steps 2 and 6. This is the Newton step. The model agrees with Problem 9, since $L(w) = \operatorname{softplus}(-w)$, and the step lowers the loss from $\ln 2 \approx 0.693$ to $L(2) = \ln(1 + e^{-2}) \approx 0.127$. In many variables it becomes $-H^{-1}\mathbf{g}$, the subject of the next page.

Where this goes wrong

1. Leaving out the n! in a coefficient

The tangent line's coefficient is just $f'(a)$, and it is natural to expect the next coefficient to be just $f''(a)$.

$f(4) = 2$, $f'(4) = \frac{1}{4}$, $f''(4) = -\frac{1}{32}$ for $f(x) = \sqrt{x}$	Right so far: the derivatives of Problem 5.

“The coefficient of $(x-4)^k$ is the k-th derivative at 4.”	The analogy that causes the mistake: the linear term, where the divisor $1! = 1$ is invisible, carried over to higher degrees.
$T_2(x) = 2 + \frac{1}{4}(x-4) - \frac{1}{32}(x-4)^2$ ✗	The $(x-4)^2$ coefficient is $f''(4)/2! = -\frac{1}{64}$: a polynomial $c_2(x-4)^2$ has second derivative $2c_2$, so matching $f''(4)$ needs the division by 2. This line gives $T_2(4.1) = 2.0246875$, whose error, 1.6×10^{-4}, is worse than that of the tangent line it was meant to improve; the correct $T_2(4.1) = 2.02484375$ is off by 1.9×10^{-6}.

2. Powers of x instead of powers of x − a

Most series met first are about 0, where $(x-a)^k$ is simply x^k.

$f(4) = 2$, $f'(4) = \frac{1}{4}$, $\frac{f''(4)}{2!} = -\frac{1}{64}$ for $f(x) = \sqrt{x}$	Right so far: the coefficients of Problem 5.
“A Taylor polynomial is the derivatives over factorials times $1, x, x^2$.”	The habit that causes the mistake: the Maclaurin form reused about $a = 4$.
$\sqrt{x} \approx 2 + \frac{1}{4}x - \frac{1}{64}x^2$ ✗	The derivatives were taken at 4, so the polynomial must be in powers of $x-4$, which are 0 there and leave $T_2(4) = f(4)$. This line fails even at the expansion point: at $x = 4$ it gives $2 + 1 - \frac{16}{64} = 2.75$, not 2. The correct polynomial is $2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2$.

3. Dropping the alternating sign of ln(1 + x)

$\dfrac{1}{1-x} = 1 + x + x^2 + \cdots$ has every sign positive, and $\dfrac{1}{1+x}$ looks almost the same.

$\ln(1+x) = \displaystyle\int_0^x \frac{dt}{1+t}$	Right so far: the route of Problem 3.

"$\dfrac{1}{1+t} = 1 + t + t^2 + \cdots$"	The analogy that causes the mistake: the geometric series read with ratio t, when $\dfrac{1}{1+t} = \dfrac{1}{1-(-t)}$ has ratio $-t$.
$\ln(1+x) = x + \dfrac{x^2}{2} + \dfrac{x^3}{3} + \cdots$ ✗	This is the series of $-\ln(1-x)$. At $x = 0.1$ it gives 0.10536, but $\ln 1.1 = 0.09531$. A concavity check catches it without a calculator: $\ln(1+x)$ is concave, so for $x > 0$ it lies below its tangent line x at 0 (Problem 1, step 5), and a series whose terms after x are all positive cannot. The correct series is $x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$.

4. Remainder bound with the derivative's maximum in the wrong place

The approximation is made at $x = 0.5$, and it seems natural to evaluate everything there.

$\lvert\ln 1.5 - T_n(0.5)\rvert = \dfrac{n!}{(n+1)!\,(1+c)^{n+1}}(0.5)^{n+1}$ for some c in $(0, 0.5)$	Right so far: Problem 6, steps 2 and 3 before the bound.
"Evaluate the derivative at the point where the function is approximated, $c = 0.5$."	The shortcut that causes the mistake: c replaced by x, as if the remainder's derivative were taken at x.
$\lvert\ln 1.5 - T_n(0.5)\rvert \le \dfrac{(1/3)^{n+1}}{n+1}$, which is below 10^{-3} from $n = 4$ ✗	c is unknown, and $(1+c)^{-(n+1)}$ is largest at $c = 0$, not at $c = 0.5$; using the smallest value makes the "bound" too small. Its answer fails: the actual error of $T_4(0.5)$ is 4.4×10^{-3}, more than four times the target. The largest value, at $c = 0$, gives $n = 7$ (Problem 6).

5. Using a series outside its radius

$\ln(1+x)$ is defined for every $x > -1$, so its series looks usable for every such x.

$\ln(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \cdots$ for $\lvert x\rvert < 1$	Right so far: Problem 3.

“The series equals $\ln(1+x)$, so to get $\ln 3$ put $x = 2$.”	The shortcut that causes the mistake: the series treated as valid wherever the function is defined, without its condition $\lvert x\rvert < 1$.
$\ln 3 = 2 - \dfrac{4}{2} + \dfrac{8}{3} - \dfrac{16}{4} + \cdots$ ✗	The radius is 1, and at $x = 2$ the terms $2^n/n$ grow without bound, so the partial sums $2, 0, 2.67, -1.33, 5.07, -5.6, \ldots$ swing further each time and never settle. Use a series whose radius covers the point: Problem 8's series at $x = 2$, inside its radius 3, gives $\ln 3 = -\ln(1 - \frac{2}{3}) = \sum_{n\geq 1} \frac{(2/3)^n}{n}$.

More at eulerpad.app. This is one of 82 free practice sets with worked solutions at eulerpad.app/practice, alongside guides to the topics behind them.

EulerPad, the app, reads your handwritten math on iPad, iPhone and Android and checks it line by line. It's free to start.

This set is also Chapter 23 of *Calculus I & II*, a 250-problem workbook with worked solutions: eulerpad.app/workbooks

eulerpad.app/practice/taylor-series-and-linear-approximation