MISC

Quadratic formula:

$$ax^2+bx+c=0 \Rightarrow x_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

Monotonically in-/decreasing:

$$\forall x.\,(f'(x)>0) \Rightarrow a>b \Leftrightarrow f(a)>f(b)$$
$$\forall x.\,(f'(x)<0) \Rightarrow a>b \Leftrightarrow f(a)<f(b)$$

RSS:

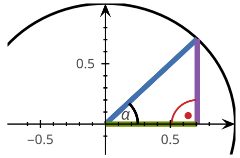
$$\mathrm{RSS}=\sum_i (y_i-f(x_i))^2$$

TRIGONOMETRY

x	0	$30°=\pi/6$	$45°=\pi/4$	$60°=\pi/3$	$90°=\pi/2$
$\sin(x)$	0	$1/2$	$\sqrt{2}/2$	$\sqrt{3}/2$	1
$\cos(x)$	1	$\sqrt{3}/2$	$\sqrt{2}/2$	$1/2$	0
$\tan(x)$	0	$1/\sqrt{3}$	1	$\sqrt{3}$	

$$\tan(x)=\frac{\sin(x)}{\cos(x)}$$
$$\sin(2x)=2\sin(x)\cos(x)$$
$$\sin^2(x)+\cos^2(x)=1$$

1, $\sin(\alpha)$, $\cos(\alpha)$



VECTORS

Term	*Definition*
Norm/Length	$\lVert v\rVert=\sqrt{\sum_{i=1}^n v_i^2}=\sqrt{v^\top v}$ $\nabla\lVert v\rVert=\frac{1}{\lVert v\rVert}v^\top$ $\nabla\frac{1}{\lVert v\rVert}=-\frac{1}{\lVert v\rVert^2}\nabla\lVert v\rVert=-\frac{1}{\lVert v\rVert^3}v^\top$ $\nabla\lVert r\rVert^2=2\lVert r\rVert\frac{1}{\lVert r\rVert}r^\top=2r^\top$
Distance	$d(u,v)=\lVert u-v\rVert$ $\nabla d(u,v)=\frac{1}{d(u,v)}((u-v)^\top(v-u)^\top)$
Scalar product	$u\cdot v=u^\top v=\sum_k u_k v_k$ $\nabla(u\cdot v)=(v^\top\ u^\top)$ $\nabla_v(u\cdot v)=v^\top$

MATRICES

Inverting:

$$A=\begin{pmatrix}a&b\\c&d\end{pmatrix}\Rightarrow A^{-1}=\frac{1}{ad-cb}\begin{pmatrix}d&-b\\-c&a\end{pmatrix}$$

Transposing:

$$\begin{pmatrix}1&0\\0&2\\3&1\end{pmatrix}^\top=\begin{pmatrix}1&0&3\\0&2&1\end{pmatrix}\quad (A^\top)^\top=A\quad (A+B)^\top=A^\top+B^\top$$
$$\begin{pmatrix}1\\4\\5\end{pmatrix}^\top=(1\ 4\ 5)\quad (\lambda A)^\top=\lambda A^\top\quad (AB)^\top=B^\top A^\top$$

Multiplication:

$$\begin{pmatrix}2&3&1\\4&-1&7\end{pmatrix}\cdot\begin{pmatrix}6&4\\1&0\\8&9\end{pmatrix}$$
$$=\begin{pmatrix}(2\cdot6+3\cdot1+1\cdot8)&(2\cdot4+3\cdot0+1\cdot9)\\(4\cdot6+-1\cdot1+7\cdot8)&(4\cdot4+-1\cdot0+7\cdot9)\end{pmatrix}=\begin{pmatrix}23&17\\79&79\end{pmatrix}$$

PROPERTIES

$$A\cdot A^{-1}=\mathbb{1}\qquad \ker A=\{v\mid Av=0\}$$
$$A+B=B+A\qquad A\cdot B\neq B\cdot A$$
$$A+(B+C)=(A+B)+C\qquad A(BC)=(AB)C$$
$$A(B+C)=AB+AC\qquad (A+B)C=AC+BC$$

DETERMINANT

$$\det(a)=a\qquad \det\begin{pmatrix}a&b\\c&d\end{pmatrix}=ad-cb$$

$$\det\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}$$
$$-a_{31}a_{22}a_{13}-a_{32}a_{23}a_{11}-a_{33}a_{21}a_{12}$$

$$\det(A\cdot B)=\det(A)\cdot\det(B)\quad \det(\mathbb{1})=1$$
$$\det(A^{-1})=\frac{1}{\det(A)}\qquad \det(A^\top)=\det(A)$$

$$\det\begin{pmatrix}0&2&1\\1&0&1\\3&4&2\end{pmatrix}=-2\det\begin{pmatrix}1&1\\3&2\end{pmatrix}+0\det\begin{pmatrix}0&1\\3&2\end{pmatrix}-4\det\begin{pmatrix}0&1\\1&1\end{pmatrix}$$
$$=-2(2-3)+0-4(-1)=2+4=6$$

$$\text{Schema}:\begin{vmatrix}+&-&+&\dots\\-&+&-&\dots\\+&-&+&\dots\\\vdots&\vdots&\vdots&\ddots\end{vmatrix}$$

EIGENVALUES

$$M\in\mathbb{R}^{n\times n},\quad M^\top=M$$
$$\det(M)=\lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_n$$

Eigenvector v for the eigenvalue λ of M:

$$Mv=\lambda v$$
$$\Leftrightarrow (M-\lambda\mathbb{1})v=0$$

Eigenvalue λ of M:

$$\det(M-\lambda\mathbb{1})=0$$

Finding eigenvalues:

$$A=\begin{pmatrix}1&1\\-2&4\end{pmatrix}$$
$$\det(A-\lambda\mathbb{1})=\det\begin{pmatrix}1-\lambda&1\\-2&4-\lambda\end{pmatrix}$$
$$=(1-\lambda)(4-\lambda)-(-2)\cdot1$$
$$=\lambda^2-5\lambda+6$$
$$\Rightarrow\lambda\in\{3,2\}$$

or alternatively

$$m=\frac12\operatorname{tr}\begin{pmatrix}1&1\\-2&4\end{pmatrix}=\frac{1+4}{2}=\frac{\lambda_1+\lambda_2}{2}$$
$$p=\det\begin{pmatrix}1&1\\-2&4\end{pmatrix}=4+2=\lambda_1\lambda_2$$
$$\lambda_{1,2}=m\pm\sqrt{m^2-p}=\frac52\pm\sqrt{\frac{25}{4}-6}$$
$$=\frac52\pm\frac12$$

INDICES

$$A\in\mathbb{R}^{n\times m},B\in\mathbb{R}^{m\times r},\qquad A\cdot B\in\mathbb{R}^{n\times r}$$
$$(A\cdot B)_{ij}=\sum_{k=1}^m A_{ik}B_{kj}\qquad=\sum_k A_{ik}B_{kj}$$
$$((A\cdot B)^\top)_{ij}=(A\cdot B)_{ji}\qquad=\sum_k A_{jk}B_{ki}$$
$$(A\cdot B^\top)_{ij}=\sum_k A_{ik}B^\top_{kj}\qquad=\sum_k A_{ik}B_{jk}$$
$$a^\top\cdot B\cdot b=\sum_{ij}(a^\top)_iB_{ij}b_j\qquad=\sum_{ij}a_iB_{ij}b_j$$
$$(B\cdot a)(C\cdot a)=\sum_i(B\cdot a)_i(C\cdot a)_i$$

ADDITION/SUBTRACTION

$$A\in\mathbb{R}^{n\times m},B\in\mathbb{R}^{n\times m}$$
$$(A+B)_{ij}=A_{ij}+B_{ij}$$
$$(A-B)_{ij}=A_{ij}-B_{ij}$$

KRONECKER DELTA

$$\delta_{ij}=(\vec e_i)_j=\begin{cases}1, & i=j\\0, & \text{otherwise}\end{cases}$$
$$(E_{ij})_{kr}=\delta_{ik}\delta_{jr}$$
$$(E_{rst})_{ijk}=\delta_{ri}\delta_{sj}\delta_{tk}$$

BASIS VECTORS ($\mathbb{R}^3$)

$$e_1=\begin{pmatrix}1\\0\\0\end{pmatrix},e_2=\begin{pmatrix}0\\1\\0\end{pmatrix},e_3=\begin{pmatrix}0\\0\\1\end{pmatrix}$$

UNIT MATRICES ($\mathbb{R}^2$)

$$E_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix},E_{12}=\begin{pmatrix}0&1\\0&0\end{pmatrix},E_{21}=\begin{pmatrix}0&0\\1&0\end{pmatrix},E_{22}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

AFFINE TRANSFORMATIONS

$$M\in\mathbb{R}^{m\times n}$$
$$A_{b,M}:\begin{cases}\mathbb{R}^m\to\mathbb{R}^n\\x\ \mapsto b+M\cdot x\end{cases}$$
$$L_A(\vec 0)=\vec 0$$

DERIVATIVES

1	$\to 0$	x	$\to 1$
x^2	$\to 2x$	x^n	$\to nx^{n-1}$
$\frac1x$	$\to -\frac{1}{x^2}$	$\sqrt{x}$	$\to \frac{1}{2\sqrt{x}}$
e^x	$\to e^x$	e^{-x}	$\to -e^{-x}$
$\ln(x)$	$\to \frac1x, x>0$	$\ln(y\cdot x)$	$\to \frac1x, x>0$
a^x	$\to \ln(a)\cdot a^x, a>0, a\neq1$	$\log_b(x)$	$\to \frac{1}{\ln(b)\cdot x}$
$\sin(x)$	$\to \cos(x)$	$\sin(2x)$	$\to 2\cos(2x)$
$\cos(x)$	$\to -\sin(x)$	$\cos(ax)$	$\to -a\sin(ax)$
$\tan(x)$	$\to 1+\tan^2(x)=\frac{1}{\cos^2(x)}$	$\operatorname{arccot}(x)$	$\to -\frac{1}{1+x^2}$
$\cot(x)$	$\to -\frac{1}{\sin^2(x)}$	$\arcsin(x)$	$\to \frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$\to -\frac{1}{\sqrt{1-x^2}}$	$\arctan(x)$	$\to \frac{1}{1+x^2}$

Case	*Rule*
Addition	$(f(x)+g(x))'=f'(x)+g'(x)$
Multiplication	$(a\cdot f(x))'=a\cdot f'(x)$
Product rule	$(f\cdot g)'=f'\cdot g+f\cdot g'$ $(f\cdot g\cdot h)'=f'\cdot g\cdot h+f\cdot g'\cdot h+f\cdot g\cdot h'$
Chain rule	$(f(g(x)))'=f'(g(x))\cdot g'(x)$
Quotient rule	$\left(\frac fg\right)'=\frac{f'g-fg'}{g^2}$

INTEGRALS

$\frac1x$	$\to \ln(\lvert x\rvert)+c$	$\ln(x)$	$\to x\ln(x)-x+c$
e^x	$\to e^x+c$	$\frac{1}{\sqrt{1-x^2}}$	$\to \arcsin(x)+c$
$\sin(x)$	$\to -\cos(x)+c$	$-\frac{1}{\sqrt{1-x^2}}$	$\to \arccos(x)+c$
$\cos(x)$	$\to \sin(x)+c$		
$\tan(x)$	$\to -\ln(\lvert\cos(x)\rvert)+c$	$\frac{1}{1+x^2}$	$\to \arctan(x)+c$
x^a	$\to \frac{1}{a+1}x^{a+1}+c, a\neq-1$		
a^x	$\to \frac{1}{\ln(a)}a^x+c$	$a^{b\cdot x}$	$\to \frac{1}{\ln(a^b)}a^x+c$
	$a^x=e^{\ln(a)\cdot x}$		

$$\arctan(x)\to x\arctan(x)-\frac12\ln(\lvert1+x^2\rvert)+c$$
$$\arccos(x)\to x\arccos(x)-\sqrt{1-x^2}+c$$
$$\arcsin(x)\to x\arcsin(x)+\sqrt{1-x^2}+c$$

RULES

$$\int f(g(x))g'(x)\,dx=F(g(x))+\text{const}$$
$$\int f(x)g(x)\,dx=F(x)g(x)-\int F(x)g'(x)\,dx$$

FUNCTIONS OF SEVERAL VARIABLES

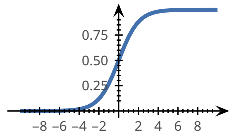
Term	*Definition*
Real valued function	$R\subset\mathbb{R}$
Vector valued function	$R\subset\mathbb{R}^m$
A function of n variables	$D\subset\mathbb{R}^n$
n-dimensional curve	$c:\mathbb{R}\to\mathbb{R}^n$
n-dimensional hyper-surface	$s:\mathbb{R}^n\to\mathbb{R}$
Reparametrization	$c(t)\to d(s)=c(h(s))$

Softmax function:

$$\sigma:\begin{cases}\mathbb{R}^n\to[0;1]^n\\x_i\ \mapsto e^{x_i}\left(\sum_{j=1}^n e^{x_j}\right)^{-1}\end{cases}$$
$$J_{ij}\sigma(x)=\delta_{ij}\frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}-\frac{e^{x_i}e^{x_j}}{\left(\sum_{k=1}^n e^{x_k}\right)^2}$$

Sigmoid function:

$$\sigma:\begin{cases}\mathbb{R}\to[0;1]\\x\mapsto\frac{1}{1+e^{-x}}=\frac{e^x}{1+e^x}\end{cases}$$
$$\sigma(-x)=1-\sigma(x)$$
$$\sigma'(x)=\sigma(x)(1-\sigma(x))$$



DERIVATIVES

Function	*Derivative*
$f:\begin{cases}\mathbb{R}\to\mathbb{R}^n\\x\mapsto y\end{cases}$	$f':\begin{cases}\mathbb{R}\to\mathbb{R}^n\\x\mapsto y\end{cases}=\begin{pmatrix}\frac{d}{dx}f_1(x)\\\frac{d}{dx}f_2(x)\\\vdots\\\frac{d}{dx}f_n(x)\end{pmatrix}$
$f:\begin{cases}\mathbb{R}^n\to\mathbb{R}\\x\mapsto y\end{cases}$	$f':\begin{cases}\mathbb{R}^n\to\mathbb{R}^{1\times n}\\x\mapsto y\end{cases}=\nabla f$ $=\left(\frac{\partial}{\partial x_1}f(x),\frac{\partial}{\partial x_2}f(x),\dots,\frac{\partial}{\partial x_n}f(x)\right)$
$f:\begin{cases}\mathbb{R}^n\to\mathbb{R}^m\\x\mapsto y\end{cases}$	$f':\begin{cases}\mathbb{R}^n\to\mathbb{R}^{m\times n}\\x\mapsto y\end{cases}=J_f=\left.\frac{\partial f}{\partial x}\right\rvert_{x=x_0}$ $=\begin{pmatrix}\nabla f_1(x)\\\vdots\\\nabla f_m(x)\end{pmatrix}=\begin{pmatrix}\frac{\partial}{\partial x_1}f_1(x)&\dots&\frac{\partial}{\partial x_n}f_1(x)\\\vdots&\ddots&\vdots\\\frac{\partial}{\partial x_1}f_m(x)&\dots&\frac{\partial}{\partial x_n}f_m(x)\end{pmatrix}$

PARTIAL VS TOTAL DERIVATIVE

$$f(x,y)=x^2+2y$$
$$\frac{\partial f}{\partial x}=2x$$
$$\frac{df}{dx}=\frac{\partial f}{\partial x}+\frac{\partial f}{\partial y}\cdot\frac{dy}{dx}=2x+2\cdot\frac{dy}{dx}\mid(\text{y dependent on x})$$
$$\text{Let } y=x^3\Rightarrow\frac{df}{dx}=2x+6x^2$$

GRADIENT VS TOTAL DERIVATIVE

$$f(x,y)=xye^y$$
$$x(t)=e^t$$
$$y(t)=t^2$$
$$\nabla f=(ye^y,xe^y+xye^y)\qquad=e^y(y,x+xy)$$
$$\frac{d}{dt}f=e^y(y,x+xy)\cdot\frac{d}{dt}\begin{pmatrix}e^t\\t^2\end{pmatrix}=\frac{d}{dt}\left(e^tt^2e^{t^2}\right)$$
$$=\left(2t+t^2+2t^3\right)e^{t^2+t}$$

GRADIENTS

Common gradients:

$$f:\mathbb{R}^n\to\mathbb{R},\ x\in\mathbb{R}^n,\ a\in\mathbb{R}^n,\ c\in\mathbb{R},\ A\in\mathbb{R}^{n\times n}$$
$$f(x)=\lVert x\rVert^2\qquad\Rightarrow\nabla f=2x^\top$$
$$f(x)=a^\top x+c\qquad\Rightarrow\nabla f=a^\top$$
$$f(x)=x^\top Ax\qquad\Rightarrow\nabla f=x^\top(A+A^\top)$$
$$f(x)=\frac12x^\top Ax+b^\top x+c\ \Rightarrow\nabla f=x^\top\frac{A+A^\top}{2}+b^\top$$
$$=x^\top A+b^\top\text{ if }A^\top=A$$
$$g:\mathbb{R}^m\to\mathbb{R},B\in\mathbb{R}^{m\times n}$$
$$f(x)=g(Bx+b)\qquad\Rightarrow\nabla f=(\nabla g(Bx+b))^\top B$$

JACOBIAN MATRIX

$$f:\mathbb{R}^n\to\mathbb{R}^m\qquad x_0\in\mathbb{R}^n$$
$$J_f(x_0):\mathbb{R}^{m\times n}\qquad J_f(x_0)_{ij}=\frac{\partial}{\partial x_j}f_i(x_0)$$
$$f:\mathbb{R}^2\to\mathbb{R}^2\quad J_f(x,y)=\begin{pmatrix}\partial x_1&\partial y_1\\\partial x_2&\partial y_2\end{pmatrix}$$

Euclid. ty lukas <3 Mat$_\mathbb{R}$

$$\begin{tikzcd}(\mathbb{R}^n,x_0)\arrow[d,"g"]\arrow[dd, bend left, "f\circ g"]\\(\mathbb{R}^m,g(x_0))\arrow[d,"f"]\\(\mathbb{R}^k,f(g(x_0)))\end{tikzcd}\quad\overset{D}{\Rightarrow}\quad\begin{tikzcd}\mathbb{R}^n\arrow[d,"J_g(x_0)"]\arrow[dd, bend left, "J_f(g(x_0))J_g(x_0)"]\\\mathbb{R}^m\arrow[d,"J_f(g(x_0))"]\\\mathbb{R}^k\end{tikzcd}$$

Chain rule:

$$f:\mathbb{R}^n\to\mathbb{R}^m,\ g:\mathbb{R}^m\to\mathbb{R}^k,\ h:\mathbb{R}^n\to\mathbb{R}^k=g\circ f$$
$$J_h=J_g(f(x))\cdot J_f(x)$$

Common jacobians:

$$f(x)=a+M\cdot x\qquad\Rightarrow J_f=M$$
$$f(x)=x\qquad\Rightarrow J_f=\mathbb{1}$$
$$f(x)=(x^\top x)x\qquad\Rightarrow J_f=2xx^\top+(x^\top x)\cdot\mathbb{1}$$
$$\Rightarrow(J_f)_{ij}=2x_jx_i+\sum_k x_k^2\delta_{ij}$$
$$f(x)=\lVert x\rVert^2\qquad\Rightarrow J_f=(2x)^\top$$
$$f_i:\begin{cases}\mathbb{R}^n\to\mathbb{R}^n\\x\ \mapsto a_i+M_ix\end{cases}$$
$$g(x)=f_n(f_{n-1}(\dots(f_1(x))))\Rightarrow J_g=M_n\cdot M_{n-1}\cdot\ldots\cdot M_1$$

LINEARISATION

$$f:\mathbb{R}^n\to\mathbb{R}^m,\ x_0\in\mathbb{R}^n$$
$$L_f(x)=f(x_0)+J_f(x_0)\cdot(x-x_0)$$

For finding better approximations for x_0:
1) Calculate J_f, then $J_f(x_0)$, then L_f
2) Find x so that $L_f(x)\overset{!}{=}0$ (Gauss)

For calculating an approximate value at x:
1) x_0 = round x to ints
2) Calculate L_f at x_0
3) Calculate $L_f(x)$ with original x

SECOND DERIVATIVE

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}(f(x,y))\right)=\frac{\partial^2f(x,y)}{\partial x^2}=\partial_{xx}f(x,y)$$

HESSIAN MATRIX

$$f:\mathbb{R}^n\to\mathbb{R},\quad x\in\mathbb{R}^n,\quad H(x)\in\mathbb{R}^{n\times n}$$
$$H(x,y)=\begin{pmatrix}f_{xx}&f_{xy}\\f_{yx}&f_{yy}\end{pmatrix}=J(\nabla f)$$
$$(H(x))_{ij}=\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}f(x)$$
$$\partial_i\partial_jf=\partial_j\partial_if$$

Common hessians:

$$f(x)=\frac12x^\top Ax+b^\top x+c\Rightarrow H_f=\frac{A+A^\top}{2}$$
$$f(x)=g(Ax+b)\qquad\Rightarrow H_f=A^\top H_g(Ax+b)A$$
$$g:\mathbb{R}^m\to\mathbb{R}$$

EXTREMAL POINTS

Quadratic form:

$$M\in\mathbb{R}^{n\times n}\qquad q_M(v)=v^\top M v\qquad v\in\mathbb{R}^n$$

$$H=\frac{1}{2}(M+M^\top)\quad \forall v\in\mathbb{R}^n.\,(q_M(v)=q_H(v))$$

$\forall v.\,(v^\top Hv>0)\vee\forall\lambda.\,(\lambda>0)\Rightarrow c_0$ is a ***local minimum***, H is ***positive definite***

$\forall v.\,(v^\top Hv\geq 0)\vee\forall\lambda.\,(\lambda\geq 0)\Rightarrow c_0$ is a ***local minimum*** or ***non-extremal***, H is ***positive semi-definite***

$\forall v.\,(v^\top Hv<0)\vee\forall\lambda.\,(\lambda<0)\Rightarrow c_0$ is a ***local maximum***, H is ***negative definite***

$\forall v.\,(v^\top Hv\leq 0)\vee\forall\lambda.\,(\lambda\leq 0)\Rightarrow c_0$ is a ***local maximum*** or ***non-extremal***, H is ***negative semi-definite***

$\forall v.\,(v^\top Hv\neq 0)\vee\forall\lambda.\,(\lambda\neq 0)\Rightarrow c_0$ is a ***saddle point***, H is ***in-definite***

Finding stationary points:

1) Calculate ∇f, then H_f
2) Find stationary point(s) x so that $\nabla f(x)\overset{!}{=}0$
3) Find type by technique of "completing squares" on $H_f(x)$ or finding its Eigenvalues

$$(v_1,v_2,v_3)\cdot\begin{pmatrix}h_{11}&h_{12}&h_{13}\\h_{21}&h_{22}&h_{23}\\h_{31}&h_{32}&h_{33}\end{pmatrix}\cdot\begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}$$

$$=v_1^2h_{11}+v_2^2h_{22}+v_3^2h_{33}+2v_1v_2h_{12}+2v_2v_3h_{23}+2v_1v_3h_{13}$$

Technique of completing squares:

$$\underbrace{x^2}_{\sqrt{\cdot}}-\underbrace{4xy}_{\div 2x}+1$$
$$=(x-2y)^2-(-2y)^2+1$$
$$=(x-2y)^2-4y^2+1$$

HYPER-SURFACES

Chain rule:

$$f:\mathbb{R}^n\to\mathbb{R}^m,\ g:\mathbb{R}^m\to\mathbb{R},\ h:\mathbb{R}^n\to\mathbb{R}=g\circ f$$

$$\nabla h(x)=\nabla g(f(x))=\nabla g(f)|_{f=f(x)}\cdot\frac{\partial}{\partial x}f(x)$$

Differentiation properties:

$$f:\mathbb{R}^n\to\mathbb{R},\ g:\mathbb{R}^n\to\mathbb{R}$$
$$\nabla(f+g)=\nabla f+\nabla g$$
$$\nabla(f-g)=\nabla f-\nabla g$$
$$\nabla(f\cdot g)=g\nabla f+f\nabla g$$
$$\nabla\frac{f}{g}=\frac{g\nabla f-f\nabla g}{g^2}$$

COMMUTATIVE DIAGRAMS

$f:D\to M$

$g:M\to R$

$h:D\to R$

$=g\circ f$

$$\begin{tikzcd} D \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] & M \arrow[r, "g"] & R \end{tikzcd}$$

GRADIENT DESCENT / NEWTON'S METHOD

GD $\quad x_{i+1}=x_i-\gamma\cdot\nabla f(x_i)$

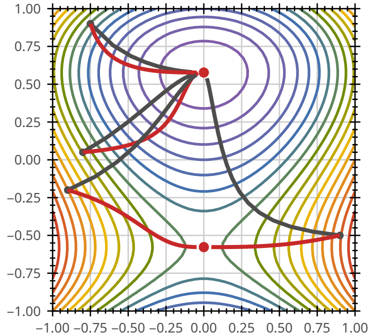
N $\quad x_{i+1}=x_i-\gamma\cdot\left(H_{f(x_i)}\right)^{-1}\cdot\nabla f(x_i)$

$\gamma=$ step size

termination criterion $\varepsilon>|x_{i+1}-x_i|$

Gradient descent doesn't always produce global minimum:

- If the function has more than one local minimum it could output that → run with different starting points
- Could output a stationary point that is not minimal (unlikely). → run with different starting points
- Could not terminate at all if step size / learning rate is so big (overshoots minimal point). Likely to happen if the graph of the function is narrow around the minimal point. → choose smaller step size



LOGISTIC REGRESSION

Term	*Definition*
Probability measure	$\mathbb{P}(E)$
Density	$f_\theta(x_i)$
Likelihood function	$L(\theta)=\prod_{i=1}^{n}f_\theta(x_i)$ e.g. $L(\theta)=\prod_{x\in S_{\text{Yes}}}p_\theta(f(x))\prod_{x\in S_{\text{No}}}1-p_\theta(f(x))$
Log-likelihood function	$\ln(L(\theta))=\sum_{i=1}^{n}\ln(f_\theta(x_i))$
Cost function (negative log-likelihood)	$c(\theta)=-\ln(L(\theta))=-\sum_{i=1}^{n}\ln(f_\theta(x_i))$
Indicator function	$\mathbb{1}_{(a;b)}(x)=\begin{cases}1 & x\in(a;b)\\0 & x\notin(a;b)\end{cases}$

$$\begin{tikzcd} \mathbb{R}^{128\times256\times3} \arrow[r, "b(x)"] \arrow[drr, "cat(x)"] & \mathbb{R} \arrow[r, "\sigma(m\cdot b-q)"] & {[0;1]} \arrow[d, "\mathbb{1}_{(t;\infty)}(\sigma)"] \\ & & \{0;1\} \end{tikzcd}$$

$b(x)$	$=$	Feature function
$\sigma(m\cdot b-q)$	$=$	Probability function
m	$=$	Weight
q	$=$	Bias
$\mathbb{1}_{(t;\infty)}(\sigma)$	$=$	Indicator function
t	$=$	Threshold

MAXIMUM LIKELIHOOD PRINCIPLE

= Finding optimal parameters θ for fitting distribution to data

1) Calculate cost function c with given distribution function f_θ, parameters θ and data points $x_1,x_2,\dots,x_n$
2) Calculate stationary point(s) of c (set ∇c to 0)
   $\Rightarrow$ Find optimum (gradient descent/newton's method if calculation not possible)

PROBABILITY THEORY

Term	*Definition*
Distributed	$\sim$
Normal distribution	$\mathcal{N}(\mu,\sigma)$
Uniform distribution	$\text{unif}(a,b)$
CDF	Cumulative Distribution Function F
PDF	Probability Density Function f

KOLMOGOROV AXIOMS

$$\mathbb{P}(\Omega)=1$$
$$\mathbb{P}(\varnothing)=0$$
$$\forall A,B\subset\Omega,\ A\cap B=\varnothing.\,(\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B))$$
$$\forall A,B\subset\Omega.\,(\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B))$$

RELATING RANDOM VARIABLES

$$\begin{tikzcd} \Omega \arrow[r, "X"] \arrow[dr, "Y"'] & \mathbb{R} \arrow[r, "F_X"] & {[0;1]} \\ & \mathbb{R} \arrow[u, "g"] \arrow[ur, "F_Y"'] & \end{tikzcd}$$

$$g(Y(\omega))=X(\omega)$$
$$\Rightarrow F_X(g(y))$$
$$=P(X(\omega)\leq g(y))$$
$$=P(g(Y(\omega))\leq g(y))$$

$$\Rightarrow\begin{cases}F_X(g(y))=P(Y(\omega)\leq y)=F_Y(y) & g\text{ increasing}\\F_X(g(y))=P(Y(\omega)\geq y)=1-F_Y(y) & g\text{ decreasing}\end{cases}$$

$$\Rightarrow\begin{cases}f_X(g(y))\cdot g'(y)=f_Y(y) & g\text{ increasing}\\f_X(g(y))\cdot g'(y)=-f_Y(y) & g\text{ decreasing}\end{cases}$$

$$|f_Y(y)\,\mathrm{d}y|=|f_X(x)\,\mathrm{d}x|\quad\text{if monotonically in/decreasing}$$

The bounds a,b of the indicator function $\mathbb{1}$ inside $f_Y(y)$ must be swapped relative to $f_X(x)$ if g is monotonically decreasing!

$$Y(\omega)=h(X(\omega))\quad \mathbb{1}_{(a;b)}\to\mathbb{1}_{(h(b);h(a))}$$

DISCRETE PROBABILITY DISTRIBUTION

Let $\Omega=\{e_1,e_2,\dots,e_n\}$ be enumerable

$$\mathbb{P}(\{e_k\})=p_k,\quad k\in\{1,2,\dots,n\},\quad E\subset\Omega$$
$$\mathbb{P}(E)=\sum_{e\in E}p(e)$$
$$F(x)=\sum_{e\in\Omega,e\leq x}p(e)$$

RULES

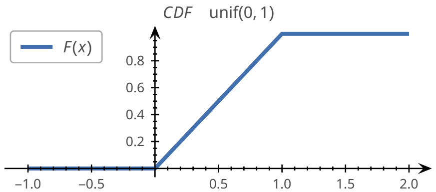
$$\forall e\in\Omega.\,(0\leq p(e)\leq 1)$$

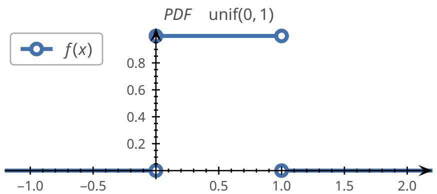
$$\sum_{e\in\Omega}p(e)=1\qquad\int_{-\infty}^{\infty}f(t)\,\mathrm{d}t=1$$

$$\lim_{x\to-\infty}F(x)=0\qquad\lim_{x\to\infty}F(x)=1$$

$$0\leq F(x)\leq 1\qquad f(t)\geq 0\text{ for }t\in\mathbb{R}$$

$F(x)$ is monotonically increasing & continuous from the right





UNIVARIATE UNIFORM DISTRIBUTION

$$X\sim\text{unif}(a,b),\quad\Omega\subset\mathbb{R}$$

$$f_X(x)=\begin{cases}\frac{1}{b-a} & \text{if }x\in(a;b)\\0 & \text{else}\end{cases}\quad=\mathbb{1}_{(a;b)}(x)\cdot\frac{1}{b-a}=F_X'$$

$$F_X(x)=\mathbb{P}((-\infty;x]\cap\Omega)\quad=\mathbb{1}_{[a;b]}(x)\cdot\frac{x-a}{b-a}+\mathbb{1}_{(b;\infty)}(x)$$

$$=\mathbb{P}(X\leq x)\quad=\int_{-\infty}^{x}f_X(t)\,\mathrm{d}t$$

UNIVARIATE NORMAL DISTRIBUTION

Standard normal distribution:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

General normal distribution:

$$f(x|\mu,\sigma)=\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

Univariate normal distribution:

$$x\sim\mathcal{N}(\mu,\sigma)$$

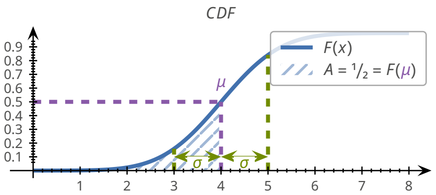
$$f(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

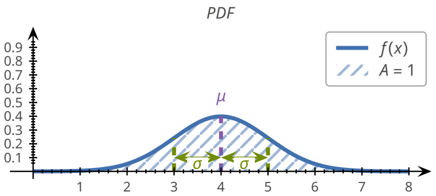
$$\mu\approx\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

$$\sigma^2\approx\text{var}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2$$

Error function:

$$\text{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$$





mean value (μ), standard deviation (σ)

MULTIVARIATE NORMAL DISTRIBUTION

$$\Sigma=\text{covariance matrix},\quad V:\Omega\to\mathbb{R}^n$$
$$V\sim\mathcal{N}(\mu,\Sigma)$$

$$f_V(v)=\frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\cdot e^{q_{\Sigma^{-1}}(v-\mu)}$$

$$q_{\Sigma^{-1}}(v-\mu)=-\frac{1}{2}(v-\mu)^\top\Sigma^{-1}(v-\mu)$$

Sample mean:

$$\mu\approx\bar{v}=\frac{1}{N}\sum_{i=1}^{N}v_i$$

Sample covariance matrix:

$$\Sigma\approx Q=\frac{1}{N-1}\sum_{i=1}^{N}(v_i-\bar{v})(v_i-\bar{v})^\top$$

STATISTICALLY INDEPENDENT

$X:\Omega\to\mathbb{R}$ and $Y:\Omega\to\mathbb{R}$ are ***statistically independent***, if the probability density f_V of the random variable

$$V:\begin{cases}\Omega\to\mathbb{R}^2\\\omega\mapsto\binom{X(\omega)}{Y(\omega)}\end{cases}$$

satisfies

$$f_V(x,y)=f_X(x)\cdot f_Y(y)$$

EXAMPLES

WORKING WITH INDICES

$$x,y\in\mathbb{R}^n,\ f:\mathbb{R}^n\to\mathbb{R}$$

$$f(x)=(x-y)^\top(3x-y)=\sum_{i=1}^{n}(x_i-y_i)(3x_i-y_i)$$

$$\partial x_i f(x)=(3x_i-y_i)+3(x_i-y_i)$$

$$f(x)=\ln(y^\top x)=\ln\left(\sum_{i=1}^{n}y_ix_i\right)\Rightarrow\nabla f=\frac{1}{y^\top x}y^\top$$

$$f:\begin{cases}\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}\\X\mapsto X\cdot A\end{cases}\qquad\frac{\partial}{\partial x_{ij}}f(X)\overset{!}{=}E_{ij}\cdot A$$

$$\frac{\partial}{\partial x_{ij}}f_{kl}(X)=\frac{\partial}{\partial x_{ij}}\sum_r x_{kr}a_{rl}=\sum_r\delta_{ik}\delta_{jr}a_{rl}=(E_{ij}\cdot A)_{kl}$$

CALCULATING LIKELIHOOD FUNCTION

Given the following data and the probability density of the form

x	*y*
1	7
2	16
3	32
4	44

$$f(t)=\lambda e^{-\lambda t}\cdot 1_{(0;\infty)}(t)$$

find a formula for the likelihood function:

$$l(\lambda)=f(7)\cdot f(16)\cdot f(32)\cdot f(44)=\lambda^4e^{-99\lambda}$$

Find a formula for the negative log-likelihood:

$$-\ln(l(\lambda))=-\ln\left(\lambda^4e^{-99\lambda}\right)=99\lambda-4\ln(\lambda)$$

CALCULATING PROBABILITY DENSITY FUNCTION

EXPONENTIAL DISTRIBUTION

Given σ, Y and X, an exponentially distributed random var

$$f_X(x)=X_{[0;\infty)}(x)\cdot e^{-x},\quad Y=\sigma(X),\quad\sigma(x)=\frac{1}{1+e^{-x}}$$

calculate the probability density function $f_Y(y)$

$$|f_Y(y)\,\mathrm{d}y|=|f_X(x)\,\mathrm{d}x|$$

$$\overset{\frac{\mathrm{d}x}{\mathrm{d}y}>0}{\Leftrightarrow}f_Y(y)=f_X(x)\cdot\frac{\mathrm{d}x}{\mathrm{d}y}$$

$$y=\frac{1}{1+e^{-x}}$$

$$\Leftrightarrow x=-\ln\left(\frac{1}{y}-1\right),\quad 0<y<1$$

$$\Rightarrow\frac{\mathrm{d}x}{\mathrm{d}y}=-\frac{1}{\frac{1}{y}-1}\cdot\frac{-1}{y^2}=\frac{1}{y-y^2}$$

$$\Rightarrow f_Y(y)=X_{[0;\infty)}(x)\cdot e^{-x}\cdot\frac{1}{y-y^2}$$

$$=\frac{1}{y^2}\cdot X_{[0;\infty)}\left(-\ln\left(\frac{1}{y}-1\right)\right)$$

$$-\ln\left(\frac{1}{y}-1\right)\geq 0\overset{y>0}{\Leftrightarrow}y\geq\frac{1}{2}$$

$$\Rightarrow X_{[0;\infty)}\left(-\ln\left(\frac{1}{y}-1\right)\right)\overset{y\in(0;1)}{=}X_{[\frac{1}{2};1)}(y)$$

$$\Rightarrow f_Y(y)=\frac{1}{y^2}X_{[\frac{1}{2};1)}(y)$$

UNIFORM DISTRIBUTION

Given $X\sim\text{unif}(0,1)$, $p>0$, $Y=X^p$ calculate the probability density function $f_Y(y)$

$$F_X(x)=\mathbb{P}(X\leq x)$$
$$F_Y(y)=\mathbb{P}(Y\leq y)\qquad=\mathbb{P}(X^p\leq y)$$
$$=\mathbb{P}\left(X\leq y^{\frac{1}{p}}\right)\qquad=F_X\left(y^{\frac{1}{p}}\right)$$
$$=\mathbb{1}_{[0;1]}(y)y^{\frac{1}{p}}+\mathbb{1}_{1;\infty}(y)$$
$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_Y(y)=\mathbb{1}_{[0;1]}(y)\frac{1}{p}y^{\frac{1}{p}-1}$$

UNIFORM DISTRIBUTION 2

$$X\sim\text{unif}(1,3),\quad Y=\sqrt[3]{\frac{5-X}{2}}$$

$$f_X(x)=\mathbb{1}_{(1;3)}(x)\frac{1}{2}$$
$$X=5-2Y^3$$
$$g(y)=5-2y^3$$
$$f_Y(y)=-f_X(g(y))g'(y)\qquad\text{monot. decr.}$$
$$=3y^2\mathbb{1}_{(1;3)}(5-2y^2)$$
$$1<5-2y^2<3$$
$$\Leftrightarrow 1<y<\sqrt[3]{2}$$
$$\Rightarrow f_Y(y)=3y^2\mathbb{1}_{(1;\sqrt[3]{2})}(y)$$