Tricks

Ch 2 Brain Teaser

idea: ① 从后往前想，每一步用前一步结论

② 从 simple case 开始，用之前的结论.

问 $n=100$，从 $n=1$ 开始想

③ 奇偶性分析！

④ 1) Indicator function when it comes to $E[X]$ 次数 2) 或 geometric dist. 3) first step

⑤ 看到 uniform $[0,1]$ 画图 $P(X_1+X_2\leq 1)$.

⑥ $P(A\cup B) = P(A)+P(B)+P(A\cup B)$

⑦ 考虑 A^c

⑧ Coin toss, cond. on (first is head)

⑨ $(n+1) \Rightarrow n$ 时 先排除 $(n+1)^{th}$，前几次找相似

⑩ Cond. on first step，下一步自己变对手！

⑪ 构造 pigeon hole problem

Ch3 Calculus & Linear Algebra

idea: ① 取log

Ch5

① 判断 mart. 用 Itô for W_t

Ch 3 Calculus & Linear Algebra
3.1 Limit & Derivatives

1. $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{u'v - v'u}{v^2}$

2. $y = f(u(x))$, $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$

3. $a^x = e^{x \ln a}$ $\quad \lim_{k \to 0} (1+x)^k = 1+kx$

$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$ $\quad \lim_{x \to 0} \frac{\sin x}{x} = 1$

4. $u = \ln y = \ln(\ln x^{\ln x}) = \ln x \cdot \ln(\ln x)$

$$\frac{du}{dx} = \frac{d \ln y}{dx} = \frac{1}{y} \cdot \frac{dy}{dx} = \frac{d \ln x}{dx} \cdot \ln(\ln x) + \frac{d(\ln(\ln x))}{dx} \cdot \ln x$$

$$= \frac{1}{x} \ln(\ln x) + \frac{d(\ln(v))}{dv} \frac{dv}{dx} \ln x$$

$$= \frac{1}{x} \ln(\ln x) + \frac{1}{v} \cdot \frac{1}{x} \ln x$$

$$= \frac{1}{x} \ln(\ln x) + \frac{1}{x \cancel{\ln x}} \cdot \cancel{\ln x}$$

$$\frac{1}{y} \cdot \frac{dy}{dx} = \frac{1}{x} \ln(\ln x) + \frac{1}{x}$$

$$\frac{dy}{dx} = \frac{\ln x^{\ln x}}{x} (\ln(\ln x) + 1)$$

5. Second derivative test

$f'(c) = 0$ & $f''(c) > 0 \Rightarrow f(x)$ has a local min at c.

6. e^π vs π^e

Assume $e^\pi > \pi^e$, $\pi \ln e > e \ln \pi$

then $\quad \frac{\ln e}{e} > \frac{\ln \pi}{\pi}$

$f(x) = \frac{\ln x}{x}$

$f'(x) = \frac{1 - \ln x}{x^2} > 0 \Rightarrow 1 > \ln x \Rightarrow x < e$

$x \geq e \quad , \quad f'(x) \leq 0 \qquad f(x) \downarrow$

$f(\pi) < f(e)$

$\frac{\ln \pi}{\ln \pi} < \frac{\ln e}{e}$

Another approach: Taylor. discuss later

7. L'Hospital's Rule

If $f(x), g(x)$ diff. $\lim_{x \to a} g'(a) \neq 0$

suppose $\lim_{x \to a} f(a) = 0$ & $\lim_{x \to a} g(a) = 0$

Or $\qquad = \pm\infty \qquad = \pm\infty$

then $\quad \lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$

(indeterminted $\Longrightarrow$ determined)

i.e. $\lim_{x \to \infty} \frac{e^x}{x^2} = \lim_{x \to \infty} \frac{e^x}{2x} = \lim_{x \to \infty} \frac{e^x}{2} = \infty$

i.e. $\lim_{x \to 0^+} x^2 \ln x$

$= \lim_{x \to 0^+} \frac{\ln x \to -\infty}{x^{-2} \to +\infty} = \lim \frac{\frac{1}{x}}{-2x^{-3}} = -\frac{1}{2}x^2 = 0$

3.2 Integration (sub, IBP, polar, normal density)

1. $f(x) = F'(x)$, $F(x)$ antiderivative $\quad \frac{dF(x)}{dx} = f(x)$

$$\int_a^b f(x) = \int_a^b F'(x)dx = [F(x)]_a^b = F(b) - F(a)$$

$$F(a) = y_a \Rightarrow F(x) = y_a + \int_a^x f(t)\,dt$$

$$\int_a^b f(x)\,dx = F(a) - F(b)$$

2. Integration by sub

$$\int f(g(x))\, g'(x)\,dx = \int f(u)\,du \quad \text{with } u = g(x) \quad du = g'(x)\,dx$$

$$\int_a^b f(g(x))\, g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du$$

别忘了!

3. IBP

$$\int u\,dv = uv - \int v\,du \qquad \int v'u\,dx = u(x)v(x) - \int u'(x)v(x)\,dx$$

① $\int \ln x\,dx = x\ln x - \int \frac{1}{x}\cdot x\,dx$

$$= x\ln x - x + C$$

② $\int_0^{\pi/6} \sec(x)\,dx = \int_0^{\pi/6} \frac{1}{\cos(x)}\,dx$

$$\frac{d\sec(x)}{dx} = \frac{d(1/\cos(x))}{dx} = \frac{\sin x}{\cos^2 x} = \frac{\tan x}{\cos x} = \tan x \sec x$$

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

$$\frac{d\,(\sec(x) + \tan(x))}{dx} = \sec x(\tan x + \sec x)$$

$$\frac{d\,\ln(\sec x + \tan x)}{dx} = \frac{1}{\quad} \sec x = \sec x$$

$$\int_0^{\frac{\pi}{6}} \sec x\,dx = \ln\left(\sec\left(\tfrac{\pi}{6}\right) + \tan\left(\tfrac{\pi}{6}\right)\right) - \ln\left(\sec(0) + \tan(0)\right)$$

$$= \ln(\sqrt{3})$$

4. Application

A.

B.

5. Expected value using integration

① $X \sim N(0,1)$, $E[X|X>0] = ?$

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

$E[X|X>0] = \frac{1}{P(x>0)}\int_0^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \overset{2x}{=} -\int_0^{-\infty} \frac{1}{\sqrt{2\pi}} e^u \, du \overset{2x}{=} -\frac{1}{\sqrt{2\pi}}[e^{-\infty} - e^0]$

$= \frac{2}{\sqrt{2\pi}}$

Let $u = -\frac{1}{2}x^2$ $du = -x\,dx$

$$E[X|X>0] = \int_{-\infty}^{\infty} x \cdot \frac{P(X=x,\, X>0)}{P(X>0)} dx$$

$$= \frac{1}{\frac{1}{2}} \int_0^\infty x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$E[X|X+Y=1]$ $X, Y \sim$ Normal $(0,1)$.

$E[X+Y|X+Y=1] = 1 = E[X|X+Y=1] + E[Y|X+Y=1]$

$E[X|X+Y] = \frac{1}{2}$

3.3 Partial Derivatives & Multiple Integrals

1. Partial Derivative.

$f(x, y)$

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{\Delta x \to 0} \frac{f(x_0+\Delta x, y_0) - f(x_0, y_0)}{\Delta x} = f_x$$

2. $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$ $\qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$

3. $f(x_1, \cdots x_n)$ $\quad x_i$ is a func of $t_1, \cdots, t_n$

$$\frac{\partial f}{\partial t_i} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial x_m}{\partial t_i}$$

4. Polar integrals

$x = r\cos\theta \qquad y = r\sin\theta$

$$\iint_R f(x, y)\,dx\,dy = \iint_R f(r\cos\theta, r\sin\theta)\, r\,dr\,d\theta$$

① $\int_0^\infty e^{\frac{-x^2}{2}}\,dx$

$$\int_0^\infty e^{-\frac{x^2}{2}}dx \int_0^\infty e^{-\frac{y^2}{2}}dy$$

$$= \int_0^\infty \int_0^\infty e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy$$

$$= \int_0^\infty \int_0^{2\pi} e^{-\frac{1}{2}r^2}\, r\,dr\,d\theta$$

$$= \int_0^\infty e^{-\frac{1}{2}r^2} r\,dr \int_0^{2\pi} d\theta$$

$$= -\int_0^{-\infty} e^u\,du \quad 2\pi$$

$$= -(0-1) \quad 2\pi$$

$$= 2\pi$$

$$\int_0^\infty e^{-\frac{x^2}{2}}dx = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}$$

$x = r\cos\theta \qquad y = r\sin\theta$

$dx = r\,dr$

$-\frac{1}{2}r^2 = u \qquad -r\,dr = du$

3.4 Important Calculus Method.

1. Taylor Series

$f(x)$ as sum of a series at $x = x_0$

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \cdots$$

☆ $x = 0$,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

n^{th} degree taylor

$$T_n(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

$f(x) = T_n(x) + R_n(x)$ → remainder

$\tilde{x}$ between x & x_0, $R_n = \frac{f^{(n+1)}(\tilde{x})}{(n+1)!} |x-x_0|^{n+1}$

let M be the maximum of $|f^{(n+1)}(\tilde{x})|$ $\forall$ $\tilde{x}$ b/w x_0 & x

$$|R_n(x)| \le \frac{M|x-x_0|^{n+1}}{(n+1)!}$$

Example i^i ?

Use taylor to prove Euler : $e^{i\theta} = \cos\theta + i\sin\theta$

$$e^{i\frac{\pi}{2}} = i$$

$$\ln(e^{i\frac{\pi}{2}}) = \ln(i)$$

$$i\frac{\pi}{2} = \ln(i)$$

$$\ln(i^i) = i\ln(i) = i^2 \frac{\pi}{2} = -\frac{\pi}{2}$$

$$i^i = e^{-\frac{\pi}{2}}$$

Example: Prove $(1+x)^n \geq 1+nx \quad \forall x>-1$ & $n\geq 2$

$f(x) = (1+x)^n$

By taylor, at $x=0$, $f(x) = 1+nx+\underbrace{\frac{\overset{+}{n}\,\overset{+}{(n-1)}\,\overset{+}{(1+\hat{x})}^{n-2}}{2!}\,\underset{+}{x}^2}_{\text{remainder}\geq 0}$

$0<\tilde{x}<x \qquad x<\hat{x}<0$

Method 2: induction.

$f(x_n) = f'(x_n)(x_{n+1}-x_n)$

2. Newton's Method. : $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

Solve $f(x)=0$ ① starting pt close to root required.

② $f(x)$ diff at root

Example : $x^2 = 37$ to 3^{rd} digit.

$f(x) = x^2 - 37$

let $x_0 = 6$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 6 - \frac{36-37}{2\times 6} = 6.083$$

$(6.083)^2 = 37.00289$

Method 2: $x_0 = 36$, taylor

$x^2 = 37$

$x = \sqrt{37}$

let $f(x) = \sqrt{x}$

$$f(x) = f(x_0) + f'(x_0)(x-x_0)$$
$$f(37) = f(36) + f'(36)(37-36)$$
$$= \sqrt{36} + \frac{1}{2}(36)^{-\frac{1}{2}}$$
$$= 6.083$$

Other Root-finding algo ?

- Bisection method
- Secant method.

3. Lagrange multiplier

Find local maximum / minimum of $f(x_1, \cdots x_n)$ s.t. $g_i(x_1 \cdots x_n) = 0$

$\nabla f = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)$ $\quad i = 1, \cdots k$

☆ $\nabla f + \lambda_1 \nabla g_1 + \cdots + \lambda_k \nabla g_k = 0$

Example: find distance from origin to $2x+3y+4z=12$

$f(x) = x^2 + y^2 + z^2$

$g_1(x) = 2x + 3y + 4z - 12 = 0$

$$\begin{cases} 2x + 2\lambda_1 = 0 \\ 2y + 3\lambda_1 = 0 \\ 2z + 4\lambda_1 = 0 \\ 2x + 3y + 4z - 12 = 0 \end{cases}$$

$D = \sqrt{x^2 + y^2 + z^2}$

4. Gradient Descent

3.5 ODE

1. Seperable DE (y, x 分开)

$$\frac{dy}{dx} = g(x)h(y)$$

i.e. $y' + 6xy = 0 \qquad y(0) = 1$

$$\frac{dy}{dx} = -6xy$$

$$\int \frac{1}{y}\,dy = \int -6x\,dx$$

$$\ln y = -3x^2 + c$$

$$y = e^{-3x^2+c}$$

$$y = e^{-3x^2} \qquad c = 0 \quad \text{by I.C.}$$

i.e. $y' = \dfrac{x-y}{x+y}$

$$(x+y)\frac{dy}{dx} = x - y$$

Let $z = x + y \qquad dz$

$$z\,\frac{d(z-x)}{dx} = x - (z-x)$$

$$z\left(\frac{dz}{dx} - 1\right) = 2x - z$$

$$\frac{z\,dz}{dx} \cancel{-z} = 2x \cancel{-z}$$

$$\int z\,dz = \int 2x\,dx$$

$$z^2 = 2x^2 + C$$

$$(x+y)^2 = 2x^2 + c$$

$$x^2 + 2xy + y^2 = 2x^2 + c$$

$$2xy + y^2 - x^2 = c$$

2. First-order linear DE

$$\frac{dy}{dx} + P(x)y = Q(x)$$ (take derivative once)

Idea: find $I(x)$ s.t. $I'(x) = I(x)P(x)$

so that $I(x)(y' + P(x)y) = I(x)Q(x)$

$$\int (I(x)\,y)' = \int I(x)\,Q(x)$$

$$y = \frac{\int I(x)Q(x)\,dx}{I(x)}$$

so $I(x) = e^{\int P(x)\,dx}$

i.e. $y' + \frac{y}{x} = \frac{1}{x^2}$ $\quad y(1) = 1$

$$I(x) = e^{\int \frac{1}{x} dx} = x$$

$$xy' + y = \frac{1}{x}$$

$$\int (yx)' = \int \frac{1}{x} dx$$

$$yx = \ln x + c$$

$$y = \frac{\ln x + c}{x}$$

$y(1) = c = 1$ $\quad$ so $y = \frac{\ln x + 1}{x}$

3. Homogenous linear equations

$$a(x)\frac{d^2y}{dx^2} + b(x)\frac{dy}{dx} + c(x)y = 0$$ (second-order DE)

a, b, c constant $\Rightarrow$ closed form solution

$$ar^2 + br + c = 0$$

① r_1, r_2 real & $r_1 \neq r_2$, $\quad y = c_1 e^{r_1 x} + c_2 e^{r_2 x}$

② r_1, r_2 real & $r_1 = r_2 = r$, $\quad y = c_1 e^{rx} + c_2 e^{rx}$

③ $r_1\ r_2$ complex $= \alpha + i\beta$ $\quad y = e^{\alpha x}(c_1\cos\beta x + c_2\sin\beta x)$

i.e. $y'' + y' + y = 0$

$b^2 - 4ac = 1 - 4 = -3 < 0$

$$r = \frac{-1 \pm \sqrt{3}\, i}{2} = \underset{\alpha}{-\frac{1}{2}} \pm \underset{\beta}{\frac{\sqrt{3}}{2}}\, i$$

$$y = e^{-\frac{1}{2}x}\left(c_1 \cos \frac{\sqrt{3}}{2}x + c_2 \sin \frac{\sqrt{3}}{2}x\right)$$

4. Non-homo eq.

$$a\frac{d^2y}{dx} + b\frac{dy}{dx} + cy = d(x)$$

先找 particular sol: y_p for -- = $d(x)$

再找 general sol: y_g for ... = 0

$$y = y_p + y_g$$

i.e. $y'' + y' + y = 1$

$$y_g = e^{-\frac{1}{2}x}\left(c_1 \cos \frac{\sqrt{3}}{2}x + c_2 \sin \frac{\sqrt{3}}{2}x\right)$$

$$y_p = 1$$

i.e. $y'' + y' + y = x$

y_g

$y_p = mx + n$

$0 + m + mx + n = x$

$m = 1 \quad n = -1$

$y_p = x - 1$

3.6 Linear Algebra

1. Vector

① Inner product

$\vec{x} \cdot \vec{y} = x^T y = \sum_{i=1}^{n} x_i y_i$ (two vectors ⇒ scalar)

② Euclidean form

$\|x\| = \sqrt{x^T x}$

$\|x-y\| = \sqrt{(x-y)^T (x-y)}$

① $\|c\vec{v}\| = c\|\vec{v}\|$

② $\|\bar{v}\| = 0 \Rightarrow \bar{v} = \bar{0}$

③ $\|\bar{u}\| + \|\bar{v}\| \geq \|\bar{u} + \bar{v}\|$

③ orthogonal : $x^T y = 0$

$\cos\theta = \dfrac{x^T y}{\|x\|\|y\|}$ (correlation coefficient)

i.e. 3 R.V. x, y, z

corr b/w x & y : $0.8 = \rho_{xy} = \cos\theta_1$

corr b/w x & z : $0.8 = \rho_{xz} = \cos\theta_2$

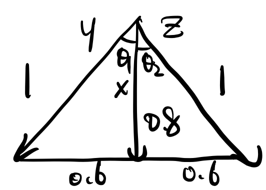
$? \leq \rho_{yz} \leq ?$

$$\frac{y^T z}{\|y\|\|z\|} = \cos(\theta_1 + \theta_2)$$

$$= 2\cos^2\theta_1 - 1$$

$$= 2 \times 0.8^2 - 1$$

$$= 0.28$$



④ Outer product $\bar{v}, \bar{w} \in \mathbb{R}^{n \times 1}$

$\bar{v} w^T$

Def linearly dep.

$\exists\ B_1 v_1 + \cdots + B_k v_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ s.t.

$B_1, \ldots B_k$ not all zero

Cols of A L.I. $\Longleftrightarrow$ A invertible

2. QR decomp

$A = QR \to$ UTM w/ positive diagonal elements

A: invertible, non-singular

Q: orthogonal matrix, $Q^{-1} = Q^T$

i.e. linear least squares regression

$y_i = \beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad \forall i = 1, \cdots, n$

$(p-1)$ regressors $\quad x_{i,0} = 1$

$Y_{n\times 1} = X_{n\times p} \beta_{p\times 1} + \varepsilon_{n\times 1}$ $\qquad \beta^T X^T X \beta$

Find $\beta = [\beta_1, \cdots, \beta_{p-1}, \beta_p]^T$

$\min_\beta f(\beta) = \min_\beta \sum_{i=1}^n \varepsilon_i^2 = \min_\beta (Y - X\beta)^T (Y - X\beta)$

$\nabla_\beta \left(\beta^T X^T X \beta - \beta^T X^T Y - Y^T X \beta + Y^T Y \right) = 0$

$2X^T X \hat\beta - X^T Y - X^T Y = 0$

$(X^T X)_{p\times p} \hat\beta = X^T Y_{p\times 1}$

Let $A = X^T X = QR$, $b = X^T Y$ (If X has LI cols) $X^T X$ invertible

$A\hat\beta = b$

so that we can perform QR on A,

$\hat\beta = (X^T X)^{-1} X^T Y$ if matrix is close to singular this will cause large numeric error

$X_{n\times p}$, $p \le n$, $\operatorname{rank}(X) = p \Rightarrow (X^T X)^{-1}$ exists

Assumption of linear least square regression

① Y & X linear: $Y = X\beta + \varepsilon$

② $E[\varepsilon_i] = 0 \quad \forall i = 1, \cdots, n$

③ $Var(\varepsilon_i) = \sigma^2 \quad \forall i = 1, \cdots, n$ constant variance

$E[\varepsilon_i \varepsilon_j] = 0 \quad i \neq j$ uncorrelated errors

④ $\rho(x_i, x_j) \neq \pm 1 \quad i \neq j$ no perfect multicolinearity ✓

⑤ ε & x_i indep. $\quad Cov(\varepsilon, X_i) = E[(\varepsilon - E\varepsilon)(X_i - EX_i)]$

$= E[\varepsilon \cdot X_i] - \cancel{E[\varepsilon]E[X_i]}$

$= E[\varepsilon \cdot X_i]$

overdetermined system $n >> p$ $\quad rank(X) = p$

$(X^T_{p \times n} X_{n \times p})_{p \times p}$ invertible

underdetermined system $n << p$ $\quad rank(X) < p$

$(X^T X)_{p \times p}$ not invertible, no unique solution

So we use ridge to find $\|\beta\|$ min!

What if $X^T X$ is not invertible? (i.e. X multicolinearity) has dep cols.

Ridge regression

$$\min \|X\beta - Y\|^2 + \lambda \|\beta\|^2$$

$$\Leftrightarrow \min \left\| \begin{bmatrix} X \\ \sqrt{\lambda} I \end{bmatrix} \beta - \begin{bmatrix} Y \\ 0 \end{bmatrix} \right\|^2$$

$$\hat{\beta} = \left(\begin{bmatrix} X \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} X \\ \sqrt{\lambda} I \end{bmatrix} \right)^{-1} \begin{bmatrix} X \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

$$= (\underbrace{X^T X + \lambda I}_{\text{always positive definite.}})^{-1} X^T Y$$

#

$v^T (X^T X + \lambda I) v$

$v^T X^T X v + \lambda \underbrace{v^T I v}_{>0}$ $\quad [I$ is PD$]$

$\underbrace{(Xv)^T X v}_{\|Xv\|^2 \geq 0} + \underbrace{\lambda v^T I v}_{>0} > 0$

$X^T X$ is PSD

3. Determinant, Eigenvalue & Eigenvector

A $n \times n$

$\det(A) = 0 \iff$ not invertible
Scalar

$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = a(ei-fh) - b(di-fg) + c(dh-eg)$$
$$= aei + bfg + cdh - afh - bdi - ceg$$

$\det(A^T) = \det(A)$ $\quad$ $\det(AB) = \det(A)\det(B)$
$\det(A^{-1}) = \frac{1}{\det(A)}$

Eigenvalue

$A \in \mathbb{R}^{n \times n}$ has a eigenvalue λ if

$\exists$ nonzero x s.t. $Ax = \lambda x$

x is called eigenvector associated w/ eigenvalue λ

Characteristic equation

linearly indep. $\rightarrow$ x can only be zero
why: otherwise $(A-\lambda I)$ invertible $(A-\lambda I)x = 0$

$\det(A - \lambda I) = 0$ $\quad$ so we can get λ

$\det(A) = \lambda_1 \cdots \lambda_n$

$\sum_{i=1}^{n} \lambda_i = \operatorname{trace}(A) = \sum_{i=1}^{n} A_{i,i}$

if n eigenvectors real & distinct

A diagonizable $\iff$ A has L.I. eigenvectors $x_1 \cdots x_n$

$\lambda_1 \cdots \lambda_n$ $\qquad$ $X = [x_1 | x_2 | \cdots | x_n]$

$$X^{-1}AX = D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$\Rightarrow$ $A = XDX^{-1}$

$\Rightarrow$ $A^k = XD^kX^{-1}$

Ex 1

$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ $\quad$ eigenvalue & vectors.

$$Ax = \lambda x$$

$$(A-\lambda I)\overset{\to \text{non-zero}}{x} = 0$$

?

$$\det(A-\lambda I) = 0$$

$$(2-\lambda)(2-\lambda)-1 = 0$$

$$4-4\lambda+\lambda^2-1 = 0$$

$$\lambda^2-4\lambda+3 = 0$$

$$\lambda_1 = 1 \quad \lambda_2 = 3 \qquad \det(A) = \lambda_1\lambda_2 = 3$$

$\lambda_1 = 1$, $\begin{cases} 2x_1 + x_2 = x_1 \\ x_1 + 2x_2 = x_2 \end{cases} \Rightarrow x_1 + x_2 = 0$

$\begin{cases} x_1 = \frac{1}{\sqrt{2}} \\ x_2 = -\frac{1}{\sqrt{2}} \end{cases}$ example

$\lambda_2 = 3$ $\begin{cases} 2x_1 + x_2 = 3x_1 \\ x_1 + 2x_2 = 3x_2 \end{cases} \Rightarrow x_1 = x_2$ $\begin{cases} x_1 = \frac{1}{\sqrt{2}} \\ x_2 = \frac{1}{\sqrt{2}} \end{cases}$ example.

$\square$

Method 2: $\det(A) = \lambda_1\lambda_2 = 4-1 = 3$

$\text{trace}(A) = 2+2 = \lambda_1 + \lambda_2$

Psuedo Inverse A^{+}

SVD

$\underset{2\times2}{A} = \underset{2\times2}{U}\,\underset{2\times2}{\Sigma}\,\underset{2\times2}{V^{T}}$ U, V orthogonal Σ diagonal ≥ 0

4. Positive semidefinite/ definite matrix

$A \in \mathbb{R}^{n\times n}$, symmetric (like covarance matrix)

all eigenvectors $\perp$ $\qquad A\vec{x} = \lambda x$

Positive Semi Definite for symmetric matrix

$\Longleftrightarrow$ ① $x^T A x \geq 0 \quad \forall\ x \in \mathbb{R}^{n\times 1}$

② all eigenvalues $\lambda \geq 0$

③ All upper left / lower right submatrices A_k, $k=1,\cdots,n$ have nonnegative determinants

corre matrix must be ↑

if no perfect L.D. b/w R.Vs

$A-\lambda I$

Positive Definite for symmetric matrix

$\Longleftrightarrow$ ① $x^T A x > 0 \quad \forall$ nonzero $x \in \mathbb{R}^{n\times 1}$

② — — > 0

③ — — positive — —

$\det(A) > 0 \neq 0 \Rightarrow$ invertible!

Ex 1 R.V. x, y, z

$corr(x,y) = 0.8$

$corr(x,z) = 0.8$

covariance matrix is semi-positivedef

b/c portfolio variance $= x^T \Sigma x \geq 0$

$$P = \begin{array}{c|ccc} & x & y & z \\ \hline x & 1 & 0.8 & 0.8 \\ y & 0.8 & 1 & \rho \\ z & 0.8 & \rho & 1 \end{array}$$

$\det(P) \geq 0$

$1-\rho^2 - 0.8(0.8-0.8\rho) + 0.8(0.8\rho - 0.8) \geq 0$

$0.28 \leq \rho \leq 1$

LU decomposition

$A \in \mathbb{R}^{n\times n}$ $\quad A = \overset{LTM}{L}\overset{UTM}{U}$ for determinant calculation

$LUx = b \Rightarrow \begin{cases} Ux = y \\ Ly = b \end{cases}$

$\det(A) = \det(L)\det(U) = \prod_{i=1}^{n} L_{ii} \prod_{i=1}^{n} U_{ii}$

if A PD (invertible),

Cholesky Decomposition $\quad \underset{P.D.}{A} = R^T \overset{UTM}{R}$ so $L = U^T$ (A is symmetric P.D. matrix)

Ex: generate two $N(0,1)$ RV w/ ρ if we have a rng of standard normal

$x_1 = z_1 \sim N(0,1)$

$x_2 = \rho z_1 + \sqrt{1-\rho^2}\, z_2 \quad \sim N(0,1)$.

$\mathrm{cov}(x_1, x_2) = \rho\, \mathrm{Var}(z_1) + 0 = \rho$

n-dimensional: Want $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1} \sim N(\mu, \overset{P.D.}{\Sigma})$ $\quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$

Let $\Sigma = R^T R$

$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ $\quad$ then $X = \mu + R^T Z \in \mathbb{R}^{n\times 1}$

SVD

$X_{n\times p} = U_{n\times p} D_{p\times p} V_{p\times p}$

cols of U span col space of X

cols of V --- row

D: diagonal matrix w/ singular value.

if X P.D. $\Rightarrow X = UDU^T$ eigen-decomposition

U: n eigenvectors; D: diagonal matrix w/ $\lambda_1 \cdots \lambda_n$

if X P.D. $\Rightarrow X_{n\times n} = U_{n\times n} D_{n\times n} U^T_{n\times n}$

$= (UD^{\frac{1}{2}})(UD^{\frac{1}{2}})^T$ $\qquad D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_p} \end{bmatrix}$

so 上题 $X = \mu + (UD^{\frac{1}{2}})Z$

Monte Carlo Simulation

1. generate RV w/ other dist. from uniform dist.

① Inversion

gen: $U \sim \text{Unif}[0,1]$

$X \sim \text{Exp}(\lambda)$.

let $F(x) = U \Rightarrow$ return $X = F^{-1}(U)$

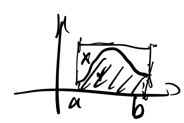
$1 - e^{-\lambda x} = U$

② Acceptance - Rejection Method.

M = upper bound, $M \geq \max_{x \in [a,b]} f(x)$

① generate $U_1 \sim U[0,1]$, $U_2 \sim U[0,1]$

② $Z_1 = a + (b-a)U_1$ $\quad Z_2 = MU_2$



③ If $Z_2 \le f(Z_1)$, return $X_0 = Z_1$

otherwise go to ①

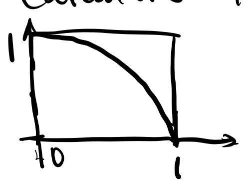
Z_1 is accepted. with prob. $P(Z_2 \le f(Z_1)) = \frac{1}{M(b-a)}$ → area → rectangular area

2. generate $[X_1 \cdots X_n]^T$ w/ $N(\mu, \Sigma)$ from $[Z_1, \cdots Z_n]^T$

where $Z_i \sim N(0,1)$

$X = \mu + R^T Z \qquad R^T R = \Sigma$ cholesky decomp.

3. Calculate π



if $Z_1^2 + Z_2^2 < 1$

accept, $a = a+1$

4. reduce var: importance sampling

find $E[h(x)]$ from x w/ $f(x)$ ✗

$E[\frac{h(x)f(x)}{g(x)}]$ from x w/ $g(x)$ ✓

Ch4 Probability Theory

4.4. Discrete & Cont. Distribution

1. Functions of R.V.

	Discrete	Cont.
cdf	$F(a) = P(X \le a)$	$F(a) = \int_{-\infty}^{a} f(x)dx$
pdf	$p(x) = P(X=x)$	$f(x) = \frac{dF(x)}{dx}$
EX	$\sum_x x\,p(x)$	$\int_{-\infty}^{\infty} x f(x)\,dx$
$E[g(x)]$	$\sum_x g(x)\,p(x)$	$\int_{-\infty}^{\infty} g(x) f(x)\,dx$
Var(X)	$E[X^2] - (EX)^2$	
Std(X)	$\sqrt{Var(X)}$	

2. Discrete RV

① Uniform $\{a, a+1, \cdots, b\}$

pmf: $p(x) = \frac{1}{a-b+1}$ $\quad x = a, \cdots b$

$EX = \frac{b+a}{2}$ $\quad Var(X) = \frac{(b-a+1)^2-1}{12}$

② Binomial

pmf: $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

$EX = np$ $\quad Var(X) = np(1-p)$

③ Poisson

pmf: $p(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$

$EX = \lambda t = Var X$

④ Geometric

pmf: $p(x) = (1-p)^{x-1} p$

$EX = \frac{1}{p}$ $\quad$ $Var X = \frac{1-p}{p^2}$

3. Conti. R.V.

① Uniform $[a, b]$

$f(x) = \frac{1}{b-a} \quad a \le x \le b$

$EX = \frac{b+a}{2} \quad Var(X) = \frac{(b-a)^2}{12}$

② Normal $N(\mu, \sigma^2)$ $\quad$ 68%, 95%, 99.7%

$\pm\sigma \quad \pm 2\sigma \quad \pm 3\sigma$

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

③ Exponential $Exp(\lambda)$

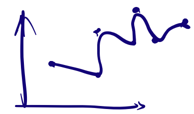
$f(x) = \lambda e^{-\lambda x}$ $\quad$ $P(X > t+s \mid X > t) = P(X > s)$

$EX = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2}$

④ Gamma

4. Concept

high bias
low var



Error = $Bias^2$ + Var + Irreducible Noise

① Bias: inability to capture true relationship — underfit — train & test error ↑

② Variance: difference in fits b/w data set — overfit — Test error ↗
high in training, low in testing

find sweet spot b/w simple & complicated models are: regularization, boosting and bagging

Min these two erros by

① Cross Validation

② Dim Reduction & feature selection

Overfitting: ① model complicated ② dimension too high (data)

Cross Validation

compare ML algo

helped with tuning param.

k-fold: a) divide data into k groups

b) fit Lasso (λ=0.2) ⇒ $\hat{\beta}$ → (x, y) c) $\hat{y}$ compare w/ y → SSR

k=4, ① ② — ⊖: train O: test

average k-testing set performance

重复 λ=0.2, 0.3 ...
get 10 SSR
λ_{min} = λ w/ SSR_{min}

sum k SSR

↑ 重复 ② ③

5. Model Evaluation

Regression

MSE: $\frac{1}{n}\Sigma(y_i-\hat{y}_i)^2$

RSS/SSE: $\Sigma(y_i-\hat{y}_i)^2$

SST (Total Sum of Squares) = $\Sigma(y_i-\bar{y})^2$ w. mean as benchmark

RMSE = $\sqrt{\frac{1}{n}\Sigma(y_i-\hat{y}_i)^2}$ (rooted)

MAE = $\frac{1}{n}\Sigma(|y_i-\hat{y}_i|)$

$R^2 = 1 - \frac{SSE}{SST}$ prop of explained y-variability (var of dep var)

$= \frac{SST - SSE}{SST} = \frac{Var(Mean) - Var(line)}{Var(mean)}$ (baseline model is $\bar{y}$)

y - x explain R^2 of variation

negative R^2 means this model is worse than predicting the mean.

not valid for non-linear models

Adjusted R^2 $\quad \bar{R}^2 = 1 - \frac{n-1}{n-p-1} R^2$ solve $R\uparrow$ as regressor $\uparrow$ no matter if it is relevant

avoid increasing features makes $R^2 \uparrow$

Classification

confusion matrix

		Predicted +	Predicted −
Actual	+	TP	FN
	−	FP	TN

Accuracy $= \frac{TP+TN}{\quad}$

Precision $= \frac{TP}{TP+FP}$ accurate of positive pre

Recall $= \frac{TP}{TP+FN}$ percentage of actual positive that are correctly identified. actual ⊕

Specificity $= \frac{TN}{TN+FP}$ actual ⊖

F1 Score $= \frac{2TP}{2TP+FP+FN}$ when class imbalanced.

3. Cross Validation

Basic idea: it measures how the model generalizes to other independent data set

Steps:

1: Divide data into K partitions (equal size)

2: Treat Fold-1 as test fold while K-1 as train folds

3: Compute **score** of test-fold

4: Repeat step 3 for all folds taking another fold as test while remaining as train

5: Take average of scores of all the folds

Perf metrics

Score of testing-fold can be customized and depends on use case, we can choose different perfo[illegible] as score. E.g. cross validation score for regression: $CV_{(k)} = \sum_{k=1}^{K} \frac{n_k}{n} MSE_k$

4. Bootstrap:

Randomly draw datasets with replacement from the training data, each sample size is the same as original training dataset. Generate B new datasets and fit model to each of these datasets.

Bootstrap can be used to estimate any aspect of the distribution of statistics derived from a model. e.g. estimate mean/variance/CI of model error.

statistically representative

Q: How to choose the value of k?

The value for k is chosen such that each train/test group of data samples is large enough to be statistically representative of the broader dataset.

A value of k=10 is very common in the field of applied machine learning, and is recommend if you are struggling to choose a value for your dataset.

If a value for k is chosen that does not evenly split the data sample, then one group will contain a remainder of the examples. It is preferable to split the data sample into k groups with the same number of samples, such that the sample of model skill scores are all equivalent.

performance good ?

Q: Difference between CV and Bootstrap

CV is used to evaluate if prediction of a model is good / is the performance of a model is good; But bootstrap is used to measure the accuracy of the estimated parameter of a model.

Frequent Q:

① How to know overfitting ?

add more training data, error ↓ significantly then.

var of model is high so ✓

② Solve overfitting

1) ↗ traing size 2) feature selection 3) regularization

Conf. Intervals

1. Conf. Level $(1-\alpha)$

   $(1-\alpha)$ of time, true value is within CI

2. Conf. Interval

$$CI_{(1-\alpha)} \quad s.t. \quad P(\theta \in CI_{(1-\alpha)}) = 1-\alpha$$

CI for mean

Distribution of X_i	Sample size n	Variance σ^2	Statistic	$1-\alpha$ confidence interval	
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known	$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$	$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
$X_i \sim$ any distribution	large	known	$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
$X_i \sim$ any distribution	large	unknown	$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\bar{X} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$	$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$
$X_i \sim \mathcal{N}(\mu, \sigma)$	no CLT small	unknown	$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$	$\left[\bar{X} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
$X_i \sim$ any distribution	small	known or unknown	Go home!		

CI for variance

Distribution of X_i	Sample size n	Mean μ	Statistic	$1-\alpha$ confidence interval
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known or unknown	$\frac{s^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$	$\left[\frac{s^2(n-1)}{\chi^2_{\frac{\alpha}{2}}}, \frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}}\right]$

Distribution of X_i	Sample size n	Mean μ	Statistic	$1-\alpha$ confidence interval
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known or unknown	$\frac{s^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$	$\left[\frac{s^2(n-1)}{\chi^2_{\frac{\alpha}{2}}}, \frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}}\right]$

Z score: # std away from μ for particular data pts

standardize the normal dist.

$$z = \frac{\bar{x}-\mu}{\sigma} \quad \text{standard normal dist.}$$

$\Updownarrow$ z table

area to the left of z score $\quad P(X<54) = P(Z<-1.22) = 0.1112$

Z score :

$\{2, 2, 3, 2, 5, 1, 6\}$ $\mu = 3$ $\sigma = 1.69$

$\frac{2-3}{1.69} = -0.59$

Hypothesis Testing: Use the ~~data~~ sample we have to determine if we can ~~support~~ reject a hypothesis made by pop. param.

Type I error: reject H_0 while H_0 true

sig level $\alpha = P(T \in R \mid H_0 \text{ true})$ T: test stat.

R: reject region

Type II error: not rej H_0 while H_0 not true

$\beta = P(T \notin R \mid H_0 \text{ not true})$

prob. of this experiment result assuming H_0 is correct

highest level of sig that H_0 cannot be rejected

p-value: prob. under H_0 of having T at least as extreme as T_0

$P(T \le T_0 \mid H_0 \text{ true})$.

$P(\quad \ge \quad)$

$P(|T| \ge |T_0| \text{ - - - - })$

prob. of getting a stat at least this far away from mean if we assume null hypo is true

ex. 

$\mu = 20$ $\mu > 20$

① H_0: $\mu = 20$ after →

H_1: $\mu > 20$

② sig level $\alpha = 0.05$

③ sample $n = 100$, $\bar{x} = 25$, s.

④ p-value: $P(\bar{x} \ge 25 \mid H_0 \text{ true})$

$= P\left(\frac{\bar{x}-\mu}{s} \ge \frac{25-20}{s}\right)$

⑤ p value $< \alpha$ rej H_0

$\ge \alpha$ ✗rej H_0

Want to test if a coin is fair

① H_0 : $P(H) = P(T) = 0.5$

H_1 : $P(H) \neq P(T)$

② sig level $= 0.05$

③ sample, toss the coin 10 times, 8H, 2T

④ p-value $:= \binom{10}{2} \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 < 0.05 =$ sig level.

⑤ we rej H_0 at 95% conf. level.

Linear Regression

$\hat\beta = \frac{Cov(x,y)}{Var(x)}$

simple: $\min \sum_{i=1}^{n} (y_i - \hat\alpha - \hat\beta x_i)^2$

w.r.t. $\hat\alpha$: $2\sum_{i=1}^{n} (y_i - \hat\alpha - \hat\beta x_i) = 0 \Rightarrow \hat\alpha = \bar y - \hat\beta \bar x$

$E(\hat\alpha) = \alpha \quad E[\hat\beta] = \beta$

w.r.t. $\hat\beta$: $-2\sum_{i=1}^{n} x_i (y_i - \hat\alpha - \hat\beta x_i) = 0$

$Var(\hat\beta) = \sigma^2 / \Sigma(x_i - \bar x)^2$

$\sum_{i=1}^{n} [x_i y_i - (\bar y - \hat\beta \bar x) x_i - \hat\beta x_i^2] = 0$

$R^2 = r^2 \quad r = cov(X, Y)$

$\sum_{i=1}^{n} (x_i y_i - \bar y x_i) + \sum_{i=1}^{n} (\hat\beta \bar x x_i - \hat\beta x_i^2) = 0$

$\hat\beta = \frac{\Sigma(x_i y_i - \bar y x_i)}{\Sigma(x_i^2 - \bar x x_i)} = \frac{E[XY] - EXEY}{Var(X)} = \frac{cov(X,Y)}{Var(X)}$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times1} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times p} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix} = X_{n\times p}\beta_{p\times1} + \varepsilon_{p\times1}$$

$\min_\beta \sum_{i=1}^{p} \varepsilon_i^2 \Rightarrow \hat\beta = (X^TX)^{-1}X^TY$ $\quad E[\hat\beta] = \beta$ unbiased

$Var(\hat\beta) = Var(\varepsilon_i)(X^TX)^{-1}$
$= \sigma^2 (X^TX)^{-1}$

Error measures:

$SSR = \sum_{i=1}^{n} \hat\varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat y_i)^2$

consistent: $n\uparrow$, $\hat\beta \to \beta$

$SSE = \sum_{i=1}^{n} (\hat y_i - \bar y)^2$ Explained

$SST = \sum_{i=1}^{n} (y_i - \bar y)^2$

$\hat\sigma^2 = Var(\hat\varepsilon_i) = \sqrt{\frac{SSR}{n-p-1}}$ Standard error for reg (ε).

$se(\hat\beta_i) = \hat\sigma \sqrt{(X^TX)^{-1}}$ standard error for $\hat\beta$

$R^2 = 1 - \frac{SSR}{SST} = \frac{SST - SSR}{SST} = \frac{Var(mean) - Var(model)}{Var(mean)}$

$y - x$ explain R^2 of variation

negative R^2 means this model is worse than predicting the mean.

not valid for non-linear models

Compare models w/ diff # indep vars

Adjusted R^2 $\quad \bar R^2 = 1 - \frac{n-1}{n-p-1} R^2$ solve $R\uparrow$ as #regressor $\uparrow$ no matter if it is relevant

Conf. Interval for β_i: $\hat\beta_i \pm t_{n-p-1,\alpha/2} \cdot SE(\hat\beta_i)$

$\hat\sigma^2 \sqrt{(X^TX)^{-1}}$

Assumption :

① x, y. linear in coeff.

② $\rho(x_i, x_j) \neq \pm 1$ $i \neq j$ no perfect colinearity

③ $E[\varepsilon_i] = 0$ $\forall\, i = 1, \cdots, n$

④ $E[\varepsilon_i \varepsilon_j] = 0$ (no heteroskedasticity) uncorrelated errors $i \neq j$
no auto correlation

$Var(\varepsilon_i) = \sigma^2$ constant var $i = 1, \cdots, n$
homo kedasticity

⑤ x_i, ε_i indep, $Cov(x_i, \varepsilon_i) = 0$

⑥ $\varepsilon_t \sim$ Normal (weak assumption)

① Violate ——— add $\log x$, $\sqrt{x}$, x^2 --- to make it linear

② Violate multicolinearity $y = \beta_a x_a + \beta_b x_b = (\beta_a + 2\beta_b) x_a$ (sys of linear eqs underdetermined)

1) $\hat{\beta}$ is sensitive to the small change in the model

2) ↓ precision of $\hat{\beta}_i$, conf. Interval not precise

can't trust the p-value

Measure: $VIF_i(x_i) = \frac{1}{1 - R_i^2}$ R_i obtained from (x_i对 $\{x_1, \cdots x_{i-1}, x_{i+1} \cdots x_k$ LR)

if co, $R_i \to 1$, $VIF \longrightarrow +\infty$

$VIF < 5$ OK

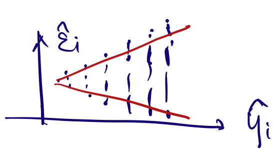
$VIF > 10$ X

Improve : 1) feature Selection.

2) 含有 domain knowledge

3) Ridge Regression

④ Violate $Var(\varepsilon_i) = \sigma^2$ residual plot



$Var(\hat{\varepsilon}_i) \nearrow$ as $\hat{y}_i \nearrow$

$Var(\beta)$ not precise $\longrightarrow$ CI for β X

Improve Box-Cox transform

$Y \sim X$

$f(Y) \sim X$

$$f(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0 \end{cases} \Rightarrow \text{corr}(\text{Var}(\hat{y}_i), \hat{\varepsilon}_i)$$

Violate $E[\varepsilon_i \varepsilon_j] \neq 0$ Heteroscedasticity

$$DW = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2} \in (0,4) \text{ usually}$$

empirically $DW = 2$, no auto correlation

$0 < DW < 1$, $\oplus$

$3 < DW < 4$, $\ominus$

Cause:

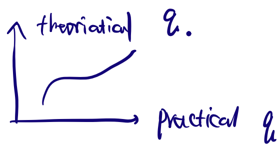
① feature 不足

② 先 $y - f(t)$ 再 LR

seasonality ⇐ autocorrelation

⑥ Normal violate

1) Q-Q plot



$df = n-p-1$

$\hat{\sigma}\sqrt{(X^TX)^{-1}}$

$Var(\hat{\beta}) = \sigma^2 / \Sigma(x_i - \bar{x})^2$

$se(\hat{\beta}) = \frac{\sigma}{\sqrt{\Sigma(x_i - \bar{x})^2}}$

Frequent Q:

① Test β_i is statistically significant ?

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$

t-statistics $t_i = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)} = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 (X^TX)^{-1}}}$

#sd that $\hat{\beta}_i$ is away from 0 !

use t-test when pop var $Var(\beta)$ unknown

→ z test is $n \to \infty$

if $t_i > t_{n-p-1, \frac{\alpha}{2}}$

or $t_i < -t_{n-p-1, \frac{\alpha}{2}}$,

rej H_0

② Test at least one variable is useful.

$H_0: \beta_1 = \beta_2 = \cdots = 0$

H_1: not all 0

$$F = \frac{(TSS - RSS)/P}{RSS/(n-p-1)}$$

($y_i - \bar{y}_i$ above TSS, $y_i - \hat{y}_i$ above RSS)

if $F > F_p$,

③ How to find $\hat{\beta}_i$ in LR? MLE, likelihood.

$L = \prod_{i=1} f(\varepsilon_i) = \dots e^{-\frac{1}{2\sigma^2}\sum \varepsilon_i^2}$ since $\varepsilon_i \sim N(0, \sigma^2)$.

④ Why least Square error? ① Diff. at $X=0$ ② more penalty on large error.

⑤ reduce effect of outliers. if not moving them

① reweights outliers ② least absolute deviation as loss function less panelty on large errors.

⑥ $\begin{pmatrix} x_1 \\ \vdots \\ x_{10} \\ x_1 \\ \vdots \\ x_{10} \end{pmatrix}$ $\begin{pmatrix} y_1 \\ \vdots \\ y_{10} \\ y_1 \\ \vdots \\ y_{10} \end{pmatrix}$ $\hat{\beta}$, t value, p-value change

$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$ p: # indep. var

$\hat{\beta} = (X^TX)^{-1}X^TY$ $E[\hat{\beta}] = \beta$ $\hat{\sigma} = \sqrt{\frac{RSS}{n-p-1}}$

$Var(\hat{\beta}) = \hat{\sigma}^2\sqrt{(X^TX)^{-1}} = \frac{\sigma^2 \text{ same}}{\sum(x_i - \bar{x})^2 \uparrow \times 2}$ $\downarrow \frac{1}{2}$

so CI for β_i $\downarrow \sqrt{2}$

$t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)\downarrow} \uparrow$ p-value $\uparrow$

Intuitively: double $\Rightarrow$ more conf. $\hat{\beta}$

t statistics $\uparrow$ since we know $\hat{\beta}$ is more sig diff. from zero

so p-value $\downarrow$ (highest level of sig that we can not rej H_0).

⑦ $y = b_1 x$ $x = b_2 y$ X, Y dataset normal. β_1 & β_2?

$y = \hat{\beta}_1 x + \varepsilon$ $\hat{\beta}_1 = \frac{cov(X,Y)}{var(X)} = \frac{\rho \cancel{\sigma_x}\sigma_y}{\cancel{\sigma_x^2}} = \rho\frac{\sigma_y}{\sigma_x}$

$x = \hat{\beta}_2 y + \varepsilon$ $\hat{\beta}_2 = \frac{cov(X,Y)}{Var(Y)} = \frac{\rho\sigma_x\cancel{\sigma_y}}{\cancel{\sigma_y^2}} = \rho\frac{\sigma_x}{\sigma_y}$

$\hat{\beta}_1\hat{\beta}_2 = \rho^2$

⑧ feature selection for linear regression

overfitting

lr.summary()

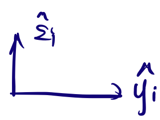
see p-value > 0.05 not stat sig t-test

⑨ How to detect outliers

an univariate outlier 不一定 regression outliers

Scatter plot $\frac{\hat{\varepsilon}_r}{std(\hat{\varepsilon}_i)}$

residual plot



standardized eliminate how large of residual is

⑩ Ridge : L_2 norm

$\min \|X\beta - y\|^2 + \lambda \|\beta\|_2 \rightarrow |\beta_1|^2 + \cdots + |\beta_n|^2$

smoothing solution β stable : for multi co : $y = (\beta_a + 2\beta_b) X_n$ more weights on β. β stable.

⑪ Lasso : L_1 norm

$\min \|X\beta - y\|^2 + \lambda \|\beta\|_1$ $\beta_i = 0$ sparse solution.

feature selection.

Time series (AR, MA, ARMA)

1. Stationary (① easy to analyze ② assumption for using tools)

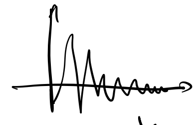
① Def

① mean of ts is constant

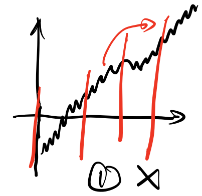
② Variance of ts is constant

③ no seasonality (periodic)

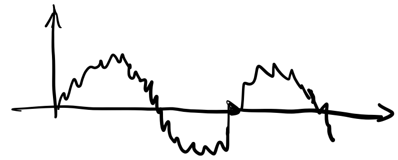
WN is stationary (mean = 0)



② X



① X



③ X

② Check for seasonality

① global vs local test for μ, σ^2

② Augmented - Dicky - Fuller Test (ADF)

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

$$\Delta y_t = \mu + \delta \underline{y_{t-1}} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

nonstationary according to H_0

$H_0: \delta = 0$ $\quad \delta = \phi - 1$ $\quad t_{\hat{\beta}_i} = \dfrac{\hat{\beta}_i}{se(\hat{\beta}_i)}$

$H_1: \delta < 0$

③ Handle stationary

1) difference: $z_t = y_t - y_{t-1}$

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

$$z_t = y_t - y_{t-1} = \beta_0 + \beta_1 t + \varepsilon_t - \beta_0 - \beta_1 (t-1) - \varepsilon_{t-1}$$

$$= \beta_1 + (\varepsilon_t - \varepsilon_{t-1}) \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$E[z_t] = \beta_1 \qquad Var(z_t) = 2\sigma^2$

Unit Root (then ts not stationary)

example = AR(1) $a_t = \phi a_{t-1} + \varepsilon_t = \phi^t a_{t-1} + \sum_{k=0}^{t-1} \phi^k \varepsilon_{t-k}$

$E[a_t] = \phi E[a_{t-1}] \cdots = \phi^t a_0$

$Var(a_t) = \sigma^2(\phi^0 + \phi^2 + \cdots + \phi^{2(t-1)})$

① $|\phi| < 1$, $E[a_t] \to 0$

$Var(a_t) = \frac{\sigma^2}{1-\phi^2}$

② $|\phi| > 1$

$E[a_t] \nearrow \infty$ exploding

③ $|\phi| = 1$

$E[a_t] = a_0$

$Var(a_t) = t\sigma^2$

改善 $d_t = a_t - a_{t-1}$

Misc

Convexity

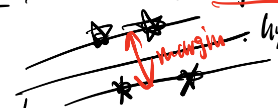
$$f(\lambda\vec{x}+(1-\lambda)\vec{y}) \le \lambda f(\vec{x})+(1-\lambda)f(\vec{y})$$

if convex programming, we can use Kuhn-Tucker cond. to find the optimal solution.

SVM

classification: binary

SVM ^ choose hyperplane which seperates the dp. of diff. classes as widely as possible, so within the margin, there is ~~free~~ little of data pts



优: ① convex optimization ② good for small dataset.

劣: ① some hyperparam (user-define)

loss function: hinge loss $\min \; \max(0, 1-\underbrace{y_i(wx_i-b)}_{\ominus \text{ if classify wrong}}) + \lambda\overbrace{|W|}^{\text{regularization}}$ penalty

kernel: C: 等价于映射到高维度方便分类.

Support Vector Regression: Instead, the margin now cover all data pts. $\underbrace{\text{soft margin}}_{\text{similar to LR}}$: allow a few dp outside of margin

优: computation friendly.

劣:

Finance

Fixed Income

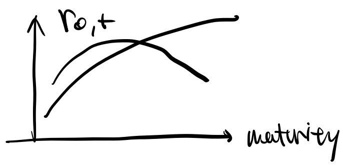
1. Yield to Maturity: anticipated return if she holds the bond until it matures.

2. Discount Bond

$$P_0 = \frac{F}{(1+R_1)(1+R_2)\cdots(1+R_T)} = \frac{F}{(1+r_{0,T})^T}$$

$R_T \equiv$ one-year spot rate of interest

$r_{0,T} \equiv$ today's T-year spot rate of interest



$$1 + r_{0,T} = \sqrt[T]{(1+R_1)\cdots(1+R_T)}$$

one-year forward rate $\quad 1 + f_t = \frac{(1+r_{0,t})^t}{(1+r_{0,t-1})^{t-1}} = \frac{P_{0,t-1}}{P_{0,t}}$

current forecast

can be t_1 & t_2

3. Coupon Bond

$$P_0 = \frac{C}{(1+y)} + \frac{C}{(1+y)^2} + \cdots + \frac{C}{(1+y)^T}$$

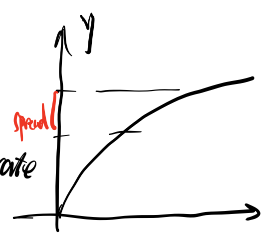
Investor Expectations

y − maturity: yield curve

(longer term bond more risks b/c more attractive investment may arise

yield spread

so longer term bond offer liquidity premium coupon rate



normal

flatten: short-term yield ↑, long term ↓

(Fed)

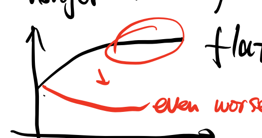
近期risk更大

They don't require risk premium for long term bond b/c they expect (no invest opp.)

$\text{Yield} = \frac{\text{Coupon}}{\text{price}\uparrow}$

more people demand long term bond

longer maturity higher yield



flatten

4. Bond Duration

P — yield

measures the sensitivity of bond price change to changes in market interest rate (how risky)

$$D_m = \sum_{k=1}^{T} k\, w_k \qquad D_m^* = \frac{D_m}{1+y} = -\frac{1}{P}\frac{\partial P}{\partial y}$$

$$w_k = \frac{C_k/(1+y)^k}{P} = \frac{PV(C_k)}{P}$$

① $T \uparrow$, Duration $\uparrow$ (4 days vs 20 years)

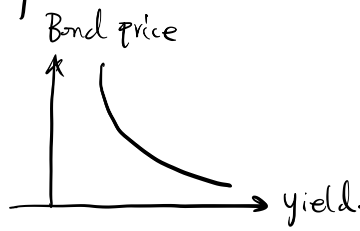
② coupon rate $\uparrow$, Duration $\downarrow$ (先得到 payment 去了)

③ YTM $\uparrow$. Duration $\downarrow$

Another interpretation: Weighted time of cash flow

how long does it take to get all money back!

5. Bond Convexity



Sensitivity of duration to yield change

$$V_m = \frac{1}{P}\frac{\partial^2 P}{\partial y^2}$$ (more curved, greater convexity)

$$P(y') = P(y)\left[1 - D_m^* \underbrace{(y'-y)}_{\text{change in yield}} + \frac{1}{2} V_m \underbrace{(y'-y)^2}_{\text{volatility in yield}}\right]$$

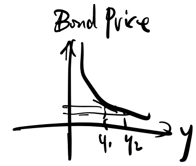
$$\frac{\Delta P}{P} \approx -D\,\Delta y + \frac{1}{2} C\,\Delta y^2$$

$$\frac{\Delta P}{P} \approx -D\,\Delta y$$ if Δy small

$T\uparrow \Rightarrow D\uparrow$ (time of getting principal back $\uparrow$, time $\uparrow$

$c\uparrow \Rightarrow D\downarrow$ (put more weights on cash flow in the near future, weighted time $\downarrow$)

$y\uparrow \Rightarrow D\downarrow$ $\quad D = -\frac{1}{P}\frac{dP}{dy}$ $\quad$ $y: 0 \to 0.1\%$
$y: 4\% \to 4.1\%$



$T\uparrow \Rightarrow C\uparrow$

$c\uparrow \Rightarrow C\downarrow$

$y\uparrow \Rightarrow C\downarrow$

X, K R.V. find K s.t. $E[|X-K|]$

$\Longleftrightarrow \underset{K}{\operatorname{argmin}} E[|X-K|^2]$

Why Fixed Income ?

Interest + capability + learn $\rightarrow$ (future)

major (finance + statistics + mathematics + programming)

① deepest largest market in financial space which help govern & corporates raise debt

② For individual investor, it helps capital stability, diversification from equity

③ fixed income is less volatile. I am a risk-averse people.

④ I like numbers & likes working in a vibrant environment

Interest rate (time value of money)

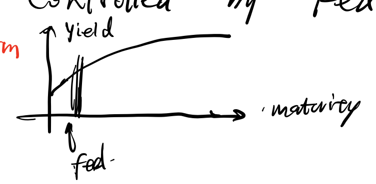
① Interest rate ↓ 2% → 1% easy money policy, economy grow

investment ↑ consumption ↑ saving ↓ $\xrightarrow{\text{long}}$ inflation ↑

↑ slow down econ ↓ $\xrightarrow{\text{long}}$ inflation ↓

② Federal Funds Rate (short term) controlled by Fed

treasury yield → little impact on long term yield.



③ Open market operation :

(10 year treasury yield high, fed buy !)

Quantitative Easing

④ interest rate ↑ Δ bond. price ↓

higher cost to issue bond

original low-yield bond not attracting, demand ↓ price ↓

⑤ Interest rate Stock

immediately 利空

一周 not sure

⑥ Positive ρ b/w Bank & interest rate

⑦ fedral funds rate ↑ 1-year treasury rate ↑ 10 year little impact

Stochastic Calculus

1. BM : cont. sto processes , $t \geq 0$ if

① $W(0) = 0$ ② increment $W(t_1) - W(0)$, $W(t_2) - W(t_1)$ indep.

③ increment $W(t_{i+1}) - W(t_i) \sim N(0, t_{i+1} - t_i)$

Property : ① $E[W(t)] = 0$, $E[W^2(t)] = t$, $E[W(t+s) | \mathcal{F}_t] = W_t$

② $cov(W(s), W(t)) = s$ $E[W^3(t)] = 0$ by symmetry

③ Markov Property

④ $Cov(B_t, B_t^2) = E[B_t^3] - E(B_t)E[B_t^2] = 0 - 0 = 0$

Ex : ① $Y_t = W^2(t) - t$

② $Z_t = \exp\{\lambda W(t) - \frac{1}{2}\lambda^2 t\}$

stopping time $\tau = \min\{t \geq 0 : W_t = 1 \text{ or } W_t = -1\}$

Itô Lemma : (like chain rule in sto calculus)

$$df(t, W_t) = f_t(t, W_t)\, dt + f_x(t, B_t)\, dB_t + \frac{1}{2} f_{xx}(t, B_t)\, dt$$

If $X(t) = X(0) + \int_0^t \beta(t, X)\, du + \int_0^t \gamma(t, X)\, dW(u)$ Itô Process

$$df(t, X_t) = f_t(t, X_t)\, dt + f_x(t, X_t)\, dX_t + \frac{1}{2} f_{xx}(t, X_t)\, dX\, dX$$

$dX\, dX = \gamma^2(t, X)\, dt$ $dX_t = \beta(t, X)\, dt + \gamma(t, X)\, dW(t)$

$$df(t, X_t) = f_t(t, X_t)\, dt + f_x(t, X_t)\, \beta(t, X)\, dt + f_x(t, X_t)\, \gamma(t, X)\, dW(t) + \frac{1}{2} f_{xx}(t, X_t)\, \gamma^2(t, X_t)\, dt$$

用于判断 $Z_t = \sqrt{t} B_t$ mart. ? (Sto Integral mart.)

Itô Isometry

$$E[(\int_0^t \Delta(u)\, dW(u))^2] = E[\int_0^t \Delta^2(u)\, du]$$

Thm: $X(t) = \int_0^t \Delta(s)\, dW_s \sim N(0, \int_0^t \Delta^2(s)\, ds)$

Option Pricing

T, $\tau = T - t$, S, r, y, σ, c, p, C, P, D, K, PV

1. price direction of options $\quad c = SN(d_1) - Ke^{-r\tau}N(d_2)$

pay off of a call: $\max(S-K, 0)$ 有权买

- ——— - —— put: $\max(K-S, 0)$ ···卖

European can only exercised at maturity. American anytime before

① $K\uparrow$, call price $\downarrow$ (buy high 3)

② $T-t$ uncertain (看

③ $S\uparrow$, call $\uparrow$ (underlying)

④ $\sigma\uparrow$, call $\uparrow$, put $\uparrow$

⑤ $r\uparrow$, call $\uparrow$ (hold cash $\uparrow$), put $\downarrow$

2. Put-call Parity

$$c + K^{-r\tau} = p + S - D$$

Coding

list : ordered, mutable

颠倒 list[::-1] 不改变

取全部 list[:]

list.insert(1, "Erdai") append

k = list.pop() 出列 last element

list.remove("xixi")

del list[1]

"abc" x 3 = "abcabcabc"

listA.extend(listB) #A, B合并

listA.append(listB) # listB 单一元素 [..., []]

listA.index("Erdai") = 1

listA.count("Erdai") = 0

listA.reverse() 颠倒, 改变list

Tuple : ordered ("a", "b", "c") immutable! but similar to list.

轻量使用, computational convenient

tupleA. count(), in, .index("a")

listA = list(tupleA) tupleA = tuple(listA)

String: immutable! ordered

listA = list(stringA)

t = stringA.split("-")

list to string: "-".join(listA) = "a-b-c"

Dictionary: unordered, key-value pair {'a':1, 'b':2} update

dictA['a']   dictA.get('a')

A = dict.keys()   list(A) = ['a', 'b']

dict.values()

dict.update{'c':3}   dict.update{'b':20}

dict['c'] = 3   dict['b'] = 20

For loop:

range(10)   [0, ..., 9]   range(0, len(list))

对 list & tuple 一样

对 dict: for each_key in dict.keys()

            print(dict.get(each_key))

While loop

Boolean: & = and | = or

int(True) = 1

continue 跳出本次 loop

break —— 所有 loop

Functions   var   type   initial value

def func(x: float = 0) → float:

    '''
    Zrdai doc string
    '''
    ...
    return anw

def get_power(a, n=2)   default 放后!

a , b = get(10)   (10, 10) tuple

Lambda Function

sq = lambda x: x*x   sq(10)   (lambda x: x*x)(10)

get_sum = lambda x,y: x+y   get_sum(10, 20)

变量 不是函数名

List Comprehension

① 函数运算

squares = [x*x for x in range(10)]

= [get_power(x, 2) for x in range(10)]

= [(lambda x: x*x)(x) for x in range(10)]

操作   param

② 逻辑运算

纵有都要

[x if x%2==0 else "odd" for x in list(range(10))]

[x for x in range(20) if x%2==0]

Map

① 函数运算

- 逻辑 + list

list(map(lambda x: x*x, range(10))) = [0, 1, ---, 81]

Filter

② 逻辑运算

用 True, False 也可!

list(filter(lambda: x%2==0, range(10)))

Reduce

from functools import reduce

reduce(lambda x, y: x+y, range(10))

把一串变成一个值

OOP #Customize class

抽象 大写 也可del

Class Student (~~object~~):

↙指代：我在initialize自己

def __init__ (self, name, age): (initialize)

① ②

#attribute

self.name = name

self.age = age

self.~~best~~_class = 'Math'

def print_age (self):

print(self.age)

def set_new_name (self, new_name):

self.name = new_name

def get_age_power (self, n)

res = self.age ** n

return res

object 具体

① ②

st = Student ('Peter', 19) st.name = 'Peter'

St.print_age() # 19

St.age = st.age + 5 # 24 更新了object的 attribute

Modules

from functools import reduce reduce ( )

import sample St = Sample.Student ('Peter', 19)

同一文件夹中的.py

Error Handling

try: b = a + 2

except:

print( )

finally:

print ("done") ← 一定执行！

Pandas.

Series

se=pd.Series([1,2,3,4]) → 0 1 2 3 index / 1 2 3 4 value (key-value pair

(np.array([1,2,3,4]))

({'a':1, 'b':2}) → a b index / 1 2 value ordered.

(------, index=[ ])

se.values se.values

se['a'] se[['a','b']] slide value

np.exp(se) → [ ]

se1 + se2 对应 index 相加, like vector

换成 np.array([ ]) or list( )

se1 + 2 每个 index 都 +2

se1%2==0 T F T T

se1[se1%2==1] slide

2 in se.values (找 value) 2 in se (找 index)

pd.isnull(se) T F T T

不同 index 相加时 不叠匀

se.index = [, , ,] 改 index 名!

se.drop(['d', 'e']) #删行

Dataframe

data = { 'state' : ['Oh,' 'WI', 'NY'],
'year' : [2002, 2019, 2022],
'pop' : [1.5, 1.7, 3.6] }

df2=pd.DataFrame (data, col=[ ],) #换 col 顺序
index=[ ] #加 index

df2[' '] df2.state # get col # series

df2.iloc[1] df2.loc['two'] # get row

NaN 无法判断 相等

df2['debt'] = [ ]   # 加一列

或   Series([ ], index=[ ])

del df2['debt']   或   df2[[ ]] 去选col

df2.T

df2.drop(['two', 'three'], axis = 0)   # 删行   fix one col, do the operation and — apply to col

=1   # 删列   iterate 每个row

index pd.DataFrame & pd.Series

obj = pd.series([ ], index=[ ])

obj[obj < 2]

df.loc['Ohio':] / .iloc[ ] / data[1:] # 选行

[: 'Ohio']

df[df > 5]   # 二维筛

df.loc[['OH', 'WI']]['one'] → series

选行 pd.df   对df选列

df[df.three > 5]

运算

se1 * se2   index 要 match

df1 + df2   index & col 都 match 否则 NaN

df1.add(df2, fill_value = 0)

df.apply(lambda x: x.max() - x.min()). → series

对每个 col 求 max − min !

回去看

排序

se.sort_values()   se.sort_index()

df. sort_values (by = 'a') a:小在前 大在后

, ascending = False

默认都是 fix one col , do the operation

se. rank (method = 'min')

每个 value 替换成对应的 ranking

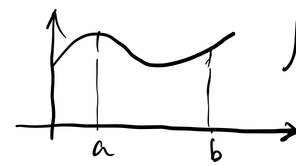
df. sum / . mean (skipna = False) / . idxmax ( ) / . cumsum ( )

最大值对应的 index

df[' ']. unique ( ) xxx in ?

Monte Carlo Simulation

1. Monte Carlo Integration in Python



$\int_a^b f(x)\,dx$

$X_i \sim \text{Unif}[a, b]$

$$\frac{1}{n}\sum_{i=1}^{n}(b-a)f(X_i) \approx \int_a^b f(x)\,dx$$

BoA QSG Superday Questions 2021

1^{st}: 中国人

① 1) time series data stationary? plot? (time series/ML)

2) 介绍给不懂的人

② mean-reverting process (Stochastic)

1) OU process (Monte Carlo)

$$dr_t = \beta(\alpha - r_t)dt + \sigma dW_t$$

α, β 含义

2) how to estimate β, α using historical data (MLE, MoM)

③ Logistic regression (ML)

1) loss function

2) interpretation.

④ Linear Programming (Calculus)

distance b/w pt to a X-Y plane

用 lagrangim multiplier

origin 到 $2x+3y+4z=12$

$$\min \quad x^2+y^2+z^2$$

$$\text{s.t.} \quad 2x+3y+4z-12=0$$

$$\begin{cases} 2x+\lambda\cdot 2=0 & x=-\lambda \\ 2y+\lambda\cdot 3=0 & y=-\frac{3}{2}\lambda \\ 2z+\lambda\cdot 4=0 & z=-2\lambda \\ 2x+3y+4z-12=0 & \end{cases}$$

$$-2\lambda+\left(-\frac{9}{2}\lambda\right)+(-8\lambda)-12=0$$

$$-\frac{29}{2}\lambda=12$$

$$\lambda=\frac{-24}{29}$$

2nd: resume + BQ 国人
leadership skill (group project)
on resume ?

3^{rd}：印度人

15 min resume

① OOP？ C++？

② BQ：group project 遇到过的问题

4^{th}：印度女

15 min resume

① $y = X\beta$

solve $\beta = (X^T X)^{-1} X^T y$

what if $X^T X$ not invertible

$(X^T X + \alpha I)$ 一定 invertible

② brain teaser

1) easy

2) 4男 4女，握别人 hundsband 手！

③ Risk-neutral measure.

组

equity → derivative
equity → cash equity algo trading