Math Foundations for ML

10-606

Geoff Gordon

Notes and reminders

- HW1 and Quiz1 are graded and viewable on Gradescope
- Further reading: Magnus & Neudecker
  - https://onlinelibrary.wiley.com/doi/book/10.1002/9781119541219
  - note their notation differs slightly from class; e.g., they define a function df(x; dx) which takes the place of our f'(x) dx

$$\ell(x) = ax + b \qquad D = \{(-2, 0), (0, -1), (1, 1)\}$$

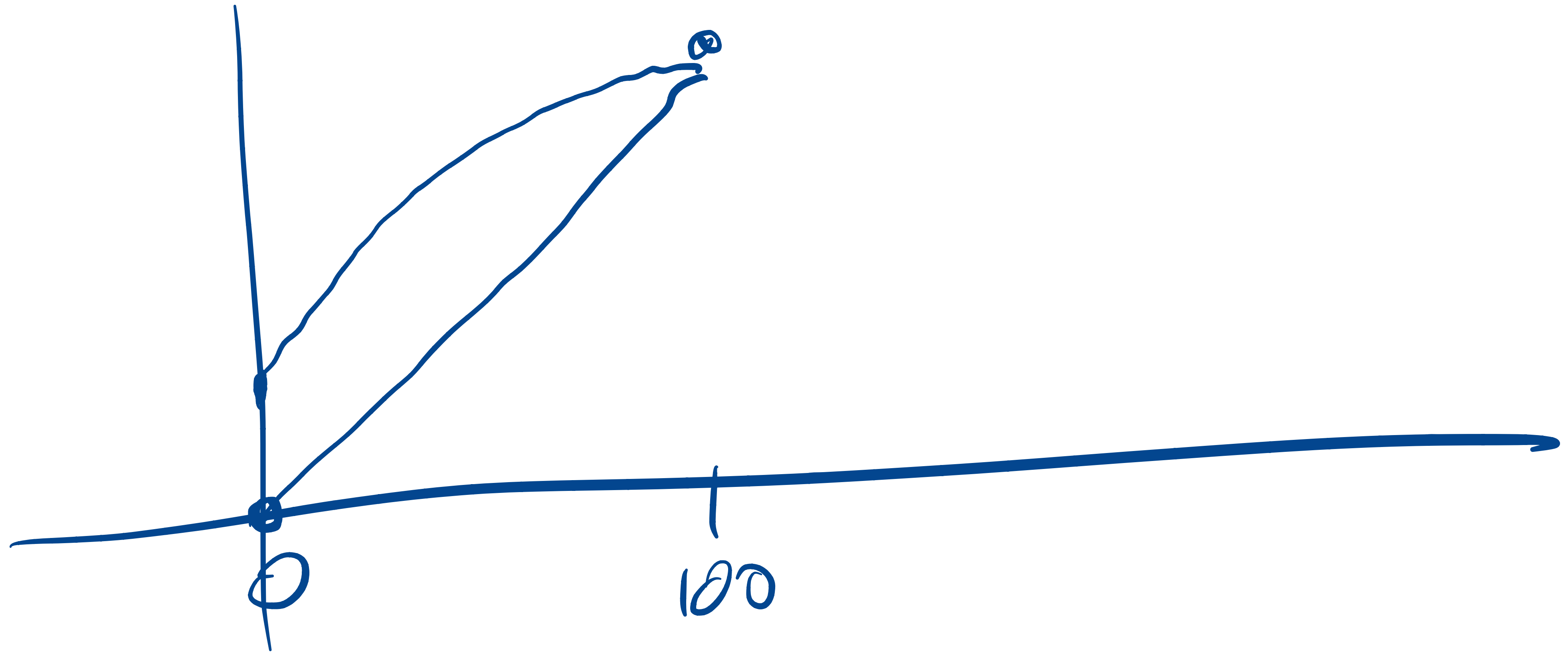
x y

$$(a(-2) + b - 0)^2 \rightarrow 4a^2 - 4ab + b^2$$

$$+ (a(0) + b - (-1))^2 \rightarrow b^2 + 2b + 1$$

$$+ (a(1) + b - 1)^2 \rightarrow a^2 + 2ab - 2a - 2b + 1$$

$$5a^2 \rightarrow 10a$$




5
0
-5
2
0
-2
-4
-6
3
2
1
0
-1
2
0
-2
2
0
-2

$$y = \ln \det A = f(\underset{\hookrightarrow \mathbb{R}^{n\times n}}{A})$$

$$\boxed{dy = \langle A^{-1}, dA \rangle}$$

$$dy = \underbrace{f'(A)}_{g}\, dA$$

$$\frac{dy}{dA}$$

$$A = f(X) = B^T X B \rightarrow \mathbb{R}^{d\times n}$$

$$\mathbb{R}^{d\times d}$$

$$\boxed{dA = \underbrace{B^T dX\, B}_{g'(x)\,dx},}$$

$$\rightarrow \langle BB^T, dX \rangle$$

$$y = f(g(x))$$

$$dy = \langle (B^T X B)^{-1}, B^T dX\, B \rangle$$

$$= f'(g(x))\, g'(x)\, dx$$

$$A = LU \qquad \det A = \det L \cdot \det U$$

$$\det L = \prod_i L_{ii}$$

$$\det U = \prod_i U_{ii}$$

$$U, V \in \mathbb{R}^{n\times n}$$

$$\langle U, V \rangle = \sum_{ij} U_{ij} V_{ij}$$

$$\underset{\in \mathbb{R}^{d\times T}}{X} = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_T \\ | & | & & | \end{pmatrix} \qquad \underset{\in \mathbb{R}^{1\times T}}{y} = \begin{pmatrix} y_1 & \cdots & y_T \end{pmatrix}$$

$$\hat{y} = w^T X \qquad w \in \mathbb{R}^{d\times 1} \qquad \varepsilon = y - \hat{y}$$

$$L = \varepsilon \cdot \varepsilon = (y - w^T X)(y - w^T X)^T$$

$$= yy^T + w^T X X^T w - 2 w^T X y^T$$

$$dL = 0 + -2\, dw^T X y^T + dw^T X X^T w + \underbrace{w^T d(X X^T w)}_{\to\, w^T X X^T dw}$$

$$= -2\, dw^T X y^T + 2\, dw^T X X^T w = 0$$

$$dw^T \underbrace{\left(-2 X y^T + 2 X X^T w\right)}_{= 0}$$

$$f(x,y) = \ldots$$

$$df(x,y) = f_x(x,y)\,dx + f_y(x,y)\,dy$$

$$f(u) = \ldots$$

$$\underbrace{df(u) = f'(u)\,du}_{\begin{pmatrix} f_x(u) & f_y(u) \end{pmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}}$$

$$u = \begin{pmatrix} x \\ y \end{pmatrix}$$

$$du = \begin{pmatrix} dx \\ dy \end{pmatrix}$$

$$y = f(\underbrace{W_2\, g(\underbrace{W_1 x + b_1}_{u_1}) + b_2}_{u_2})$$

$x \in \mathbb{R}^k$

$y \in \mathbb{R}^d$

$W_1 \in \mathbb{R}^{d' \times k}$ $\quad b_1 \in \mathbb{R}^{d'}$

$W_2 \in \mathbb{R}^{d \times d'}$ $\quad b_2 \in \mathbb{R}^d$

$$f(z) = \begin{pmatrix} \vdots \\ 1/1+e^{-z_i} \\ \vdots \end{pmatrix}, \quad z \in \mathbb{R}^d$$

$$g(w) = \begin{pmatrix} \vdots \\ 1/1+e^{-w_i} \\ \vdots \end{pmatrix}, \quad w \in \mathbb{R}^{d'}$$

$$y = f(u_2)$$

$$dy = f'(u_2)\, du_2 \qquad u_2 = W_2\, g(u_1) + b_2$$

$$= f'(u_2)[W_2\, dg(u_1)]$$

$$= f'(u_2)[W_2\, g'(u_1)\, du_1]$$

$$= f'(u_2)[W_2\, g'(u_1) W_1\, dx$$

$$= f'(u_2) W_2\, g'(u_1) W_1$$

$$= f'(u_2)$$

$$f(z) = \begin{pmatrix} \vdots \\ 1/(1+e^{-z_i}) \\ \vdots \end{pmatrix}$$

$$f'(z) = \begin{pmatrix} \ddots & & 0 \\ & f_i(z)(1-f_i(z)) & \\ 0 & & \ddots \end{pmatrix}$$

Jacobian of componentwise fn
is diagonal

$$X = \begin{pmatrix} | & | & | & | & | \end{pmatrix}$$

$$Y = X - \underbrace{\mu(X)}_{\text{row-wise mean}} e^T \qquad e = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$Z = \underbrace{V(Y)}_{\text{row-wise variance}}{}^{-\frac{1}{2}} \circ Y$$

$$\mu(X) = \frac{1}{n} X e$$

$$V(Y) = \frac{1}{n} \operatorname{diag}(Y Y^T)$$

$$\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 12 & 15 & 18 \\ -14 & -16 & -18 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ -2 & -2 & -2 \end{pmatrix}$$