Math Foundations for ML

10-606

Geoff Gordon

Notes and reminders

- HW2 should be released by tomorrow
  - We will extend due date for the programming part
- This week, Friday is a class; following Monday is a lab


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

$\boxed{X \in \mathbb{R}^{m \times T} \qquad y \in \mathbb{R}^{1 \times T}}$ data

$\boxed{W \in \mathbb{R}^{d \times m}}$ — params

$$\begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_T \\ | & | & & | \end{pmatrix}$$

1st layer

$$s(WX) = \tilde{u} \in \mathbb{R}^{d \times T}$$

$$ds(z) = r(z) \odot dz$$

layer 1.5

$$U = \tilde{u} / \|\tilde{u}\| \in \mathbb{R}^{d \times T}$$

$$\|\tilde{u}\| = \sqrt{\sum_{ij} \tilde{u}_{ij}^2}$$

layer 2

$$\hat{y} = vU$$

$\boxed{v \in \mathbb{R}^{1 \times d}}$

$$\hat{y} \in R^{1 \times T}$$

output

$$\boxed{L} = \|\hat{y} - y\|^2$$

$$L = \hat{y} \cdot \hat{y} - 2\hat{y} \cdot y + y \cdot y$$

$$dL = \hat{y} \cdot d\hat{y} + d\hat{y} \cdot \hat{y} - 2\, d\hat{y} \cdot y = 2(\hat{y} - y)\, d\hat{y}^T$$

$$d\hat{y} = dv\, U \quad \text{taking } \tfrac{\partial}{\partial v} \qquad dL = \boxed{2(\hat{y}-y)\, U^T}\, dv^T$$

$$d\hat{y} = v\, dU \quad \text{taking } \partial/\partial w$$

$$dU = \frac{1}{\|\tilde{u}\|}\left(d\tilde{u} - U\, tr(U^T d\hat{u})\right)$$

$$d\hat{u} = r(Wx) \circ d(Wx) = r(Wx) \circ (dW\, X)$$

$$dL = 2(\hat{y}-y)\, dU^T v^T = 2v\, dU\, (\hat{y}-y)^T$$

$$= 2v\left(\frac{1}{\|\tilde{u}\|}\left(d\tilde{u} - U\, tr(U^T d\hat{u})\right)\right)(\hat{y}-y)^T$$

$$tr(AB) = tr(BA)$$

$$\sum_i [AB]_{ii} = \sum_i \sum_j A_{ij} B_{ji} = \sum_j [BA]_{jj} = tr(BA)$$

$$= \frac{2}{\|\tilde{u}\|} \operatorname{tr}\left(v \left(d\tilde{u} - u \operatorname{tr}(u^T d\tilde{u}) \right) (\hat{y} - y)^T \right)$$

$$= \frac{2}{\|\tilde{u}\|} \left[\operatorname{tr}\left((\hat{y} - y)^T v \, d\tilde{u} \right) - \operatorname{tr}\left(v u (\hat{y} - y)^T \right) \operatorname{tr}(u^T d\tilde{u}) \right]$$

$$= \underbrace{\frac{2}{\|\tilde{u}\|} \operatorname{tr}\left(\left[(\hat{y} - y)^T v - u^T \operatorname{tr}\left(v u (\hat{y} - y)^T \right) \right] d\tilde{u} \right)}_{G^T}$$

$$dL = \operatorname{tr}\left(\boxed{G^T} d\tilde{u} \right)$$

derivative of L as fn of $\tilde{u}$

$$= \operatorname{tr}\left(G^T \left(r(WX) \circ (dW \, X) \right) \right)$$

$$= tr\left(\left(G \circ r(WX) \right)^T (dW\,X) \right)$$

$$= tr\left(\boxed{X \left(G \circ r(WX) \right)^T} dW \right)$$

derivative

$$\underbrace{m \times T \quad (d \times T)^T}$$

$m \times d \qquad d \times m$

$m \times m$

$d(A^T) = (dA)^T \qquad d\,\mathrm{tr}(A) = \mathrm{tr}(dA)$

$d\,\mathrm{numpy.reshape}(A, \ldots) = \mathrm{numpy.reshape}(dA, \ldots)$

$\mathrm{tr}(AB) = \mathrm{tr}(BA) \qquad a = \mathrm{tr}(A)$ if $a \in \mathbb{R}^{1\times 1}$

$a, s, t \in \mathbb{R}$

$u, v \in \mathbb{R}^{10}$

$M, N \in \mathbb{R}^{10\times 10}$

$ds = a\,dt \qquad ds = u^T dv \qquad ds = \mathrm{tr}(M^T dN)$

$du = v\,dt \qquad du = M\,dv$

$dM = N\,ds$

Second order

$$y \approx f(x) + df(x) + d^2 f(x)$$

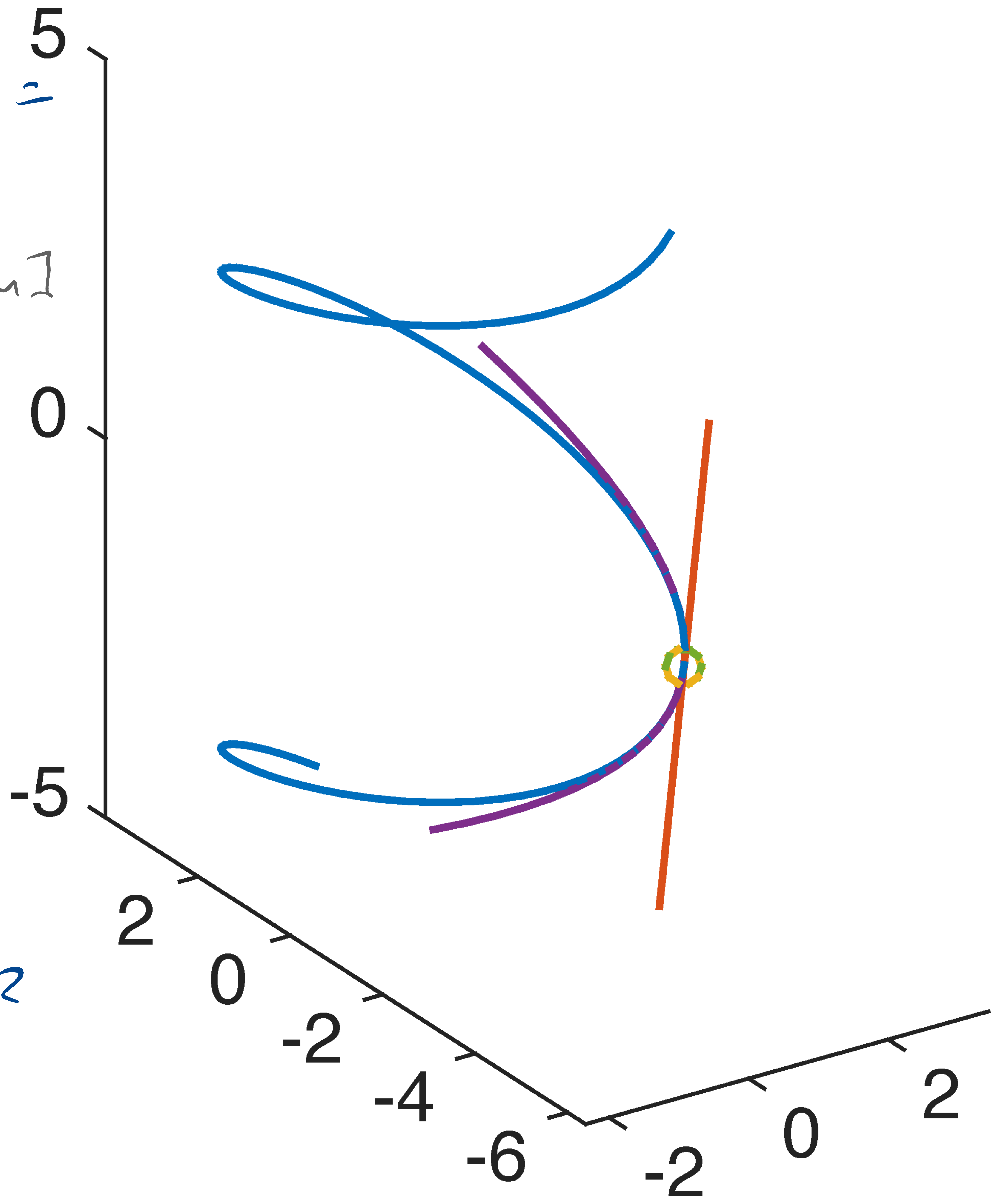
[edit: clearer in this notation]

$$f(t) = \begin{pmatrix} 2\cos t \\ 4 \sin t \\ t \end{pmatrix}$$

$$df(t) = f'(t)\,dt = \begin{pmatrix} -2\sin t \\ 4\cos t \\ 1 \end{pmatrix} dt$$

$$d^2 f(t) = f''(t)\,dt^2 = \begin{pmatrix} -2\cos t \\ 4\sin t \\ 0 \end{pmatrix} dt^2$$

$dy =$



$s = f(t)$ $\quad s, t$ scalar $\quad f''(t) \in \mathbb{R}$

$s = f(t)$ $\quad t$ scalar, s any $\quad f''(t)$ same shape as s

$s = f(t)$ $\quad s$ scalar, t vector $\quad f''(t)$ matrix

$$d^2 s = dt^T f''(t)\, dt$$

(Hessian)

$$\frac{1}{2}(Ax-b)^T(Ax-b) = \frac{1}{2}x^TA^TAx - x^TA^Tb + \frac{1}{2}b^Tb$$

$$dg = \underbrace{\frac{1}{2}dx^TA^TAx + \frac{1}{2}x^TA^TA\,dx}_{dx^TA^TAx} - dx^TAb$$

$$d^2) \quad = \quad dx^T A^T A \, dx$$