

- $df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz$ (total derivative)

$$f = f(x, y, z)$$

I Regression (not classification)

- Recall: linear model

$$W^T X = \sum w_i x_i \rightarrow \begin{matrix} \text{homogeneous} \\ \text{additivity} \end{matrix}$$

- $x_n \in \mathbb{R}^d$: feature sample for a sample

$$y_n \in \mathbb{R}: \text{observed output}$$

simple: find a function $h(x) = W^T x$ to approximate y

- measure error $(h(x) - y)^2$ (cost function)

$$\Delta \text{ training error: } E_{\text{train}}(h) = \frac{1}{N} \sum_{n=1}^N (h(x_n) - y_n)^2$$

$$\ggg w = A \setminus x$$

$$\Delta \|W^T x - y\|_2^2$$

$$W^* = (X^T X)^{-1} X^T y \quad (\text{closed-form solution})$$

\hookrightarrow invertible / non-invertible

(often $d > N$)

\downarrow
pseudo-inverse

also: complexity sparse / dense

2 - Logistic Regression

→ Binary Classification

• input : $x_1, x_2, \dots, x_n \in \mathbb{R}^d$

output : $y_1, y_2, \dots, y_n \in \{+1, -1\}$

a. training : take the sign of some value function

$$\text{sign}(f(x_i)) \approx y_i$$

b. using logistic hypothesis

$$P(y=1|x) = \theta(w^T x) \quad \text{where}$$

$$\text{where } \theta(s) = \frac{1}{1+e^{-s}}$$

$$P(y=1|x) + P(y=-1|x) = 1$$

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

\downarrow $\theta(w^T x)$ \downarrow $\theta(-w^T x)$

question: difference between likelihood & probability

check: monotonic increase

linear model / non-linear :

depends on
how you combine
your weights &
features

linearity : homogeneous
additivity

c. hinge loss

- Empirical Risk Minimization

△ $f_W(x)$: decision function to be learned

W is the parameters

△ General empirical risk minimization

$$\underset{W}{\text{minimize}} \quad \frac{1}{N} \sum_{n=1}^N \text{loss}(f_W(x_n), y_n)$$