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

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

I. ~~Regression~~ (not classification)

- Recall: linear model

  $W^T x = \sum w_i x_i$ $\rightarrow$ homogeoneous additivity

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

  $y_n \in R$: observed output

  simply: find a function $h(x) = w^T x$ to approximate y

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

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

  >>> w = A \ x

  Δ $\|W^T x - y\|_2^2$

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

  $(X^T X)$ → inveritable / non-invertable (often $d > N$

  ↓

  pseudo-inverse

  also: complexity sparse / dense

2. Logistic Regression

* Binary Classification

- input : $X_1, X_2, \dots X_n \in R^d$

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

a : training : take the sign of some value function

$$sing(f(x_i)) \approx y_i$$

b : using logistic hypothesis

$P(y=1|x) = \theta(w^T x)$ where

where $\theta(s) = \frac{1}{1+e^{-s}}$ $\quad P(y=1|x) + P(y=-1|x) = 1$

$$1 - \underbrace{\frac{e^{w^T x}}{1+e^{w^T x}}}_{\theta(W^T x)} = \frac{1+e^{w^T x}}{1+e^{w^T x}} - \frac{e^{w^T x}}{1+e^{w^T x}} = \underbrace{\frac{1}{1+e^{w^T x}}}_{\theta(-w^T x)}$$

question : difference between likelihood & probability

check : monotonic increase

linear model / non-linear : $\boxed{\text{depends on how you combine your weights \& features}}$

linearity : homo-geneous additivity

c. hinge loss

- Empirical Risk Minimization
  - △ $f_W(x)$: decision function to be learned
    W is the parameters
  - △ General empirical risk minimization

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