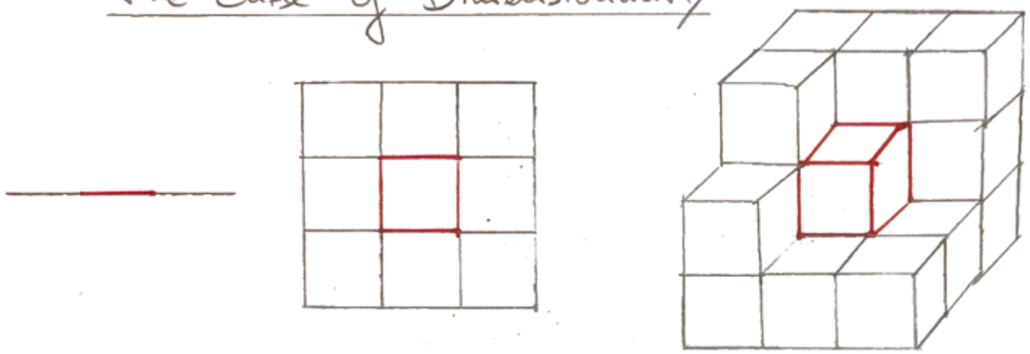
[1]

MSA220 / MVE440

Lecture 1, 2019-03-25

The Curse of Dimensionality



1D 2D 3D

Relative importance of — to all parts

$1/1$ $1/9$ $1/27$

Example of a probability mass function (pmf)

Bernoulli distribution with parameter $\Theta \in (0,1)$.

$$p(0) = \Theta \quad , \quad p(1) = 1 - \Theta$$

Example of a probability density function (pdf)

Multivariate normal distribution with mean vector $\mu \in \mathbb{R}^p$ & covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$

$$p(x) = |2\pi\Sigma|^{-\frac{p}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

Statistical decision theory for regression

Define $J(f) = \mathbb{E}_{p(x)}\left[\mathbb{E}_{p(y|x)}\left[L(y, f(x))\right]\right]$

where $L(y, f(x)) = (y - f(x))^2$.

We want to find $\hat{f}(x)$ such that

$$\hat{f} = \arg\min_f J(f)$$

The outer expectation averages over the x-values, so we can focus on minimizing

$$\mathbb{E}_{p(y|x)}\left[(y - f(x))^2\right]$$

for every x (i.e. expectation over y conditional on a fixed x). It holds that

$$\mathbb{E}_{p(y|x)}\left[(y-f(x))^2\right] = \int (y - f(x))^2 p(y|x)\, dy =$$

$$= \int (y \overbrace{- \mathbb{E}_{p(y|x)}[y] + \mathbb{E}_{p(y|x)}[y]}^{=0} - f(x))^2 p(y|x)\, dy =$$

$$= \int (y - \mathbb{E}_{p(y|x)}[y])^2 p(y|x)\, dy$$

$$+ 2\int (y - \mathbb{E}_{p(y|x)}[y])\underbrace{(\mathbb{E}_{p(y|x)}[y] - f(x))}_{\text{constant w.r.t. } y} p(y|x)\, dy$$

$$+ \int \underbrace{(\mathbb{E}_{p(y|x)}[y] - f(x))^2}_{\cancel{\text{indepen}}\ \text{constant w.r.t. } y} p(y|x)\, dy =$$

$$= \mathrm{Var}_{p(y|x)}[y] + (\mathbb{E}_{p(y|x)}[y] - f(x))^2$$

Minimal for $f(x) = \mathbb{E}_{p(y|x)}[y]$

$\boxed{2}$ **Statistical decision theory for classification**

Note that 0-1 loss can be written as

$$L(i, c(x)) = \begin{cases} 0 & i = c(x) \\ 1 & i \neq c(x) \end{cases} = \underbrace{\mathbb{1}}_{\text{indicator function}}(i \neq c(x))$$

Then

$$\mathbb{E}_{p(i|x)}\left[L(i, c(x))\right] = \mathbb{E}_{p(i|x)}\left[\mathbb{1}(i \neq c(x))\right] =$$

$$= \sum_{i=1}^{k} \mathbb{1}(i \neq c(x))\, p(i|x) = \sum_{i \neq c(x)} p(i|x) =$$

$$= 1 - p(c(x)|x)$$

This is minimal if $\hat{c}(x) = \underset{1 \leq i \leq k}{\arg\max}\; p(i|x)$.