University of Amsterdam
Informatics Institute

AMLAB
Amsterdam
Machine Learning Lab

Machine Learning 1

Lecture 2.1 - Expectation - Variance

Erik Bekkers

(Bishop 1.2.2)

Image credit: Kirillm | Getty Images

Slide credits: Patrick Forré and Rianne van den Berg

Expectations

$x \sim p(X)$

- random variable $x \in X$ and function $f : X \to \mathbb{R}$

$$\mathbb{E}[f] = \mathbb{E}_{x \sim p(X)}[f(x)] = \begin{cases} \sum_X f(x)\, p(x) & \text{- discrete} \\ \int_X f(x)\, p(x)\, dx & \text{- continuous} \end{cases}$$

$\{x_1, x_2, \ldots, x_N\}$

- For N points drawn from p(X):

$$\mathbb{E}[f] \approx \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

- Conditional expectation:

$$\mathbb{E}[f \,|\, y] = \mathbb{E}_{x \sim p(X|Y=y)}[f(x)] = \begin{cases} \sum_X f(x)\, p(x \,|\, Y=y) \\ \int_X f(x)\, p(x \,|\, Y=y)\, dx \end{cases}$$

Variance

$$\mathbb{E}[f(x)+g(x)] = \mathbb{E}[f(x)] + \mathbb{E}[g(x)]$$

$$\mathbb{E}[c\, f(x)] = \sum_x c\, f(x) p(x) = c\, \mathbb{E}[f(x)]$$

$$\mathbb{E}[c] = c$$

- The expected quadratic distance between f and its mean $\mathbb{E}[f]$

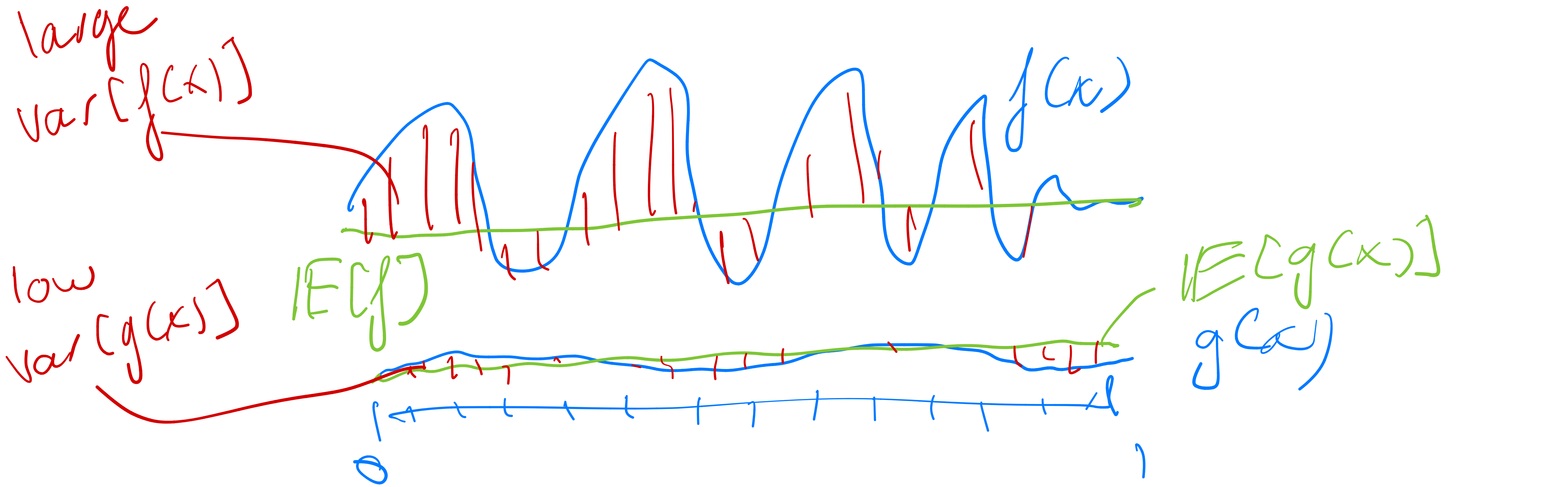
$$\mathrm{var}[f] = \mathbb{E}\left[\left(f(x) - \mathbb{E}[f(x)]\right)^2\right]$$

$$= \mathbb{E}\left[f(x)^2 - 2 f(x)\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2\right]$$

$$= \mathbb{E}[f(x)^2] - 2\mathbb{E}[f(x)]^2 + \mathbb{E}[f(x)]^2$$

$$= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$x \sim U[0,1]$ $\quad$ $f(x),\ f: X \rightarrow \mathbb{R}$



Covariance between 2 random variables

- Measures the extent to which *X* and *Y* vary together

$$\mathrm{cov}[x, y] = \mathbb{E}_{x,y \sim p(x,y)}\left[(x - \mathbb{E}[x])(y - \mathbb{E}[y])\right]$$

$$= \mathbb{E}[xy - x\mathbb{E}[y] - y\mathbb{E}[x] + \mathbb{E}[x]\mathbb{E}[y]]$$

$$= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$$

- Vectors of random variables **x** and **y**, covariance matrix:

$x, y \in \mathbb{R}^D$

$$\mathrm{cov}[\mathbf{x}, \mathbf{y}] = \mathbb{E}_{x,y \sim p(x,y)}\left[\underbrace{(x - \mathbb{E}[x])}_{\mathbb{R}^{D\times 1}}\underbrace{(y - \mathbb{E}[y])^T}_{\mathbb{R}^{1\times D}}\right] \in \mathbb{R}^{D\times D}$$

$$= \mathbb{E}[xy^T] - \mathbb{E}[x]\mathbb{E}[y]^T$$

Covariance between 2 random variables

- Covariance between independent variables: $p(x,y)=p(x)p(y)$

$\mathrm{cov}[x,y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] = 0$

$$\int_x\int_y xy\,p(x,y)\,dx\,dy = \int_x\int_y xy\,p(x)p(y)\,dx\,dy$$

$$= \int_x x\,p(x)\,dx \int_y y\,p(y)\,dy = \mathbb{E}[x]\mathbb{E}[y]$$

- Note: $\mathrm{cov}[x,y] = 0$ does not imply x,y independent

$x \sim U[-1,1] \iff \forall x\in(-1,1): p(x) = \frac{1}{2}$

$y = x^2$

$\mathrm{cov}[x,y] = \cancel{\mathbb{E}[xy]} - \cancel{\mathbb{E}[x]}\mathbb{E}[y] = 0$

"0" "because odd" "0"

$$\int_{-1}^{1} x^3 \tfrac{1}{2}\,dx = 0$$

- $\mathrm{cov}[\mathbf{x}] \equiv \mathrm{cov}[\mathbf{x},\mathbf{x}]$