

Why

1

Definition

We want to estimate a random variable $x: \Omega \to \mathbb{R}^n$ from a random variable $y: \Omega \to \mathbb{R}^n$ using an estimator $\phi: \mathbb{R}^m \to \mathbb{R}^n$ which is affine.² In other words, $\phi(\xi) = A\xi + b$ for some $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. We will use the mean squared error cost.

We want to find A and b to minimize

$$\mathbf{E}||Ax + b - y||^2.$$

Proof. Express
$$\mathbf{E}(\|Ax + b - y\|^2)$$
 as $\mathbf{E}((Ax + b - y)^{\top}(Ax + b - y))$
+ $\operatorname{tr}(A\mathbf{E}(xx^{\top})A^{\top})$ + $\mathbf{E}(x)^{\top}A^{\top}b$ - $\operatorname{tr}(A^{\top}\mathbf{E}(yx^{\top}))$
+ $b^{\top}A\mathbf{E}(x)$ + $b^{\top}b$ - $b^{\top}\mathbf{E}(y)$
- $\operatorname{tr}(A\mathbf{E}(xy^{\top}))$ - $\mathbf{E}(y)^{\top}b$ + $\mathbf{E}(yy^{\top})$

The gradients with respect to b are

so $2A\mathbf{E}(x) + 2b - 2\mathbf{E}(y)$. The gradients with respect to A are

so $2\mathbf{E}(xx^{\top})A^{\top} + 2\mathbf{E}(x)b^{\top} - 2\mathbf{E}(xy^{\top})$. We want A and b solutions to

$$A\mathbf{E}(x) + b - \mathbf{E}(y) = 0$$

$$\mathbf{E}(xx^{\top})A^{\top} + \mathbf{E}(x)b^{\top} - \mathbf{E}(xy^{\top}) = 0$$

¹Future editions will include an account.

²Actually, the development flips this. Future editions will correct.

so first get $b = \mathbf{E}(y) - A\mathbf{E}(x)$. Then express

$$\begin{split} \mathbf{E}(xx^\top)A^\top + \mathbf{E}(x)(\mathbf{E}(y) - A\mathbf{E}(x))^\top - \mathbf{E}(xy^\top) &= 0. \\ \mathbf{E}(xx^\top)A^\top + \mathbf{E}(x)\mathbf{E}(y)^\top - \mathbf{E}(x)\mathbf{E}(x)^\top A^\top - \mathbf{E}(xy^\top) &= 0. \\ (\mathbf{E}(xx^\top) - \mathbf{E}(x)\mathbf{E}(x)^\top)A^\top &= \mathbf{E}(xy^\top) - \mathbf{E}(x)\mathbf{E}(y)^\top. \\ \mathbf{cov}(x, x)A^\top &= \mathbf{cov}(x, y). \end{split}$$

So $A^{\top} = \mathbf{cov}(x, x)^{-1}\mathbf{cov}(x, y)$ means $A = \mathbf{cov}(y, x)\mathbf{cov}(x, x)^{-1}$ is a solution. Then $b = \mathbf{E}(y) - \mathbf{cov}(y, x)\mathbf{cov}(x, x)^{-1}\mathbf{E}(x)$. So to summarize, the estimator $\phi(x) = Ax + b$ is

$$cov(y, x) (cov x, x)^{-1} x + E(y) - cov(y, x) cov(x, x)^{-1} E(x)$$

or

$$\mathbf{E}(y) + \mathbf{cov}(y, x) \left(\mathbf{cov}x, x\right)^{-1} \left(x - \mathbf{E}(x)\right)$$

