Instrumental-Variables Method

Section 3.8 - Linear Regression in the Presence of Noise

Kiran Bagale

July 2025

Motivation

- Least Squares and Bayesian methods yield same solution for $\boldsymbol{w}$ in noiseless Gaussian environments.
- Assumes:
  - Input regressor $\boldsymbol{x}$ is noiseless
  - Desired response d is noiseless
- But in practice: input is noisy $\rightarrow \boldsymbol{z}_i = \boldsymbol{x}_i + \boldsymbol{v}_i$

Noisy Regression Case

- Using noisy regressor $\boldsymbol{z}$:

$$\hat{\boldsymbol{w}} = \hat{\boldsymbol{R}}_{zz}^{-1} \hat{\boldsymbol{r}}_{dz}$$

- With white noise vector $\boldsymbol{v} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$:

$$\hat{\boldsymbol{R}}_{zz} = \hat{\boldsymbol{R}}_{xx} + \sigma^2 \boldsymbol{I}, \quad \hat{\boldsymbol{r}}_{dz} = \hat{\boldsymbol{r}}_{dx}$$

- Maximum Likelihood Estimator becomes:

$$\hat{\boldsymbol{w}}_{ML} = (\hat{\boldsymbol{R}}_{xx} + \sigma^2 \boldsymbol{I})^{-1} \hat{\boldsymbol{r}}_{dx}$$

Noise as Regularizer

- Additive noise $\Rightarrow$ stabilizes solution
- Identical to MAP estimate with regularization:

$$\lambda = \sigma^2$$

- Irony: Noise acts as regularizer
- Downside: introduces bias

Need for Unbiased Estimation

- Goal: Obtain **asymptotically unbiased** estimator of $\boldsymbol{w}$
- Solution: **Instrumental-Variables Method (IVM)**
- Introduce vector $\hat{\boldsymbol{x}}$ (instrumental variable)
- Must satisfy:
    1. $\mathbb{E}[x_j \hat{x}_k] \neq 0$ (correlated with $\boldsymbol{x}$)
    2. $\mathbb{E}[v_j \hat{x}_k] = 0$ (uncorrelated with noise)

Instrumental Variables Estimation

- Compute cross-correlation matrices:

$$\hat{\boldsymbol{R}}_{z\hat{x}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{x}}_i \boldsymbol{z}_i^T, \quad \hat{\boldsymbol{r}}_{d\hat{x}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{x}}_i d_i$$

- Estimate:

$$\hat{\boldsymbol{w}}^{(N)} = \hat{\boldsymbol{R}}_{z\hat{x}}^{-1}\hat{\boldsymbol{r}}_{d\hat{x}}$$

- As $N \to \infty$, estimator is unbiased

Comparison of Methods

- **Least Squares / ML**:
  - Biased in presence of noise
  - Stabilized by regularization
- **Instrumental Variables**:
  - Asymptotically unbiased
  - Requires valid instruments satisfying properties

Choosing Instrumental Variables

- Key challenge: how to construct $\hat{\boldsymbol{x}}$
- In time-series analysis: tractable via lagged variables, etc.
- Satisfying:
  - High correlation with $\boldsymbol{x}$
  - Independence from noise $\boldsymbol{v}$

Conclusion

- Instrumental-Variables Method enables unbiased estimation
- Practical for real-world problems with noisy input
- Trade-off: complexity in generating valid instruments