Hands-On Lab Syllabus

Time Series Forecasting: From VAR to Transformers and Diffusion

Format. Six self-contained exercises in `exercises.py`, run on a CPU with only `numpy` and `matplotlib`. Each exercise has a worked part that runs as-is and `# TODO` extensions for participants. Allow ∼25 minutes guided plus optional take-home stretch goals. Figures are regenerated by `make_figures.py`.

Setup.

- Core: `pip install numpy matplotlib`; then `python exercises.py`.
- Stretch goals: `pip install chronos-forecasting gluonts torch statsforecast`.

1 Exercise 1 — Build (or load) the dataset

Goal. Create the synthetic hourly "load" series (trend + daily + weekly + noise) and hold out the last $H = 48$ hours.
Skills. Train/test splitting; recognizing seasonal structure.
You implement. Swap in a real CSV via `pd.read_csv`; find the dominant period with an autocorrelation plot.
Check. The held-out window is never touched during fitting.

2 Exercise 2 — Classical VAR(1)

Goal. Simulate a bivariate VAR(1) with a known A, recover it by OLS, and read off impulse responses $B^{(h)} = A^h$.
Skills. VAR estimation, impulse-response analysis, Granger causality.
You implement. The Granger F-test of H_0: lags of y_1 do not enter the y_2 equation; convert F to a p-value.
Check. $\hat{A} \approx A$; and $F_{y_1 \to y_2} \approx 0$ while $F_{y_2 \to y_1} \gg 0$, recovering the asymmetry built into A.
Connects to. The "From VAR to Attention" bridge — fixed A_i versus data-dependent attention weights.

3 Exercise 3 — Baselines and MASE

Goal. Build a seasonal-naive forecast with an 80% interval derived from a *rolling-origin backtest*, and score it with MASE.
Skills. Honest interval construction; scale-free point metrics.
You implement. An AutoETS baseline (`statsforecast`); compare MASE. Explain why a one-step-error interval would be too narrow.
Check. Seasonal-naive MASE $\approx$ 1.06; the band comes from backtest residuals, not in-sample one-step errors.

4 Exercise 4 — Self-attention from scratch

Goal. Implement scaled dot-product attention in ∼12 lines of numpy and inspect the attention matrix over patch tokens.
Skills. Q/K/V projections, softmax, the $\sqrt{d_k}$ scaling.

You implement. A causal mask ($S_{ij} = -\infty$ for $j > i$); a multi-head split and concatenate.
Check. Each row of the attention matrix sums to 1; with the mask the matrix is lower-triangular.

5 Exercise 5 — Diffusion forward process

Goal. Noise a window with the closed form $x^{(n)} = \sqrt{\bar{\alpha}_n}\, x^{(0)} + \sqrt{1 - \bar{\alpha}_n}\, \epsilon$ at several levels n.
Skills. Forward diffusion, the $\alpha_n/\bar{\alpha}_n$ schedule.
You implement. One reverse step $x^{(n-1)} = \frac{1}{\sqrt{\alpha_n}}\big(x^{(n)} - \frac{\beta_n}{\sqrt{1-\bar{\alpha}_n}}\epsilon_\theta\big) + \sqrt{\beta_n}\, z$, then sample from pure noise (with a mock or trained ϵ_θ).
Check. At $n = N$ the window is $\approx \mathcal{N}(0, I)$ (std ≈ 1).

6 Exercise 6 — Probabilistic forecast and scoring

Goal. Generate correlated sample paths, summarize them as a quantile fan, and score with sample-CRPS and interval coverage.
Skills. Scenario forecasting, proper scoring rules, calibration.
You implement. Diagnose why an 80% band covers only ∼70% and recalibrate; as a stretch goal, replace the sample paths with Chronos-2 zero-shot draws and beat the CRPS.
Check. CRPS ≈ 2.3 and coverage ≈ 0.7 for the starter model.

Mapping to the course modules

Exercise	Module	Core idea practiced
1	1–5	Data, held-out evaluation
2	1 (Classical)	VAR, Granger causality, impulse response
3	2 (Transformers)	Baselines, MASE, honest intervals
4	2 (Transformers)	Self-attention mechanics
5	4 (Diffusion)	Forward/reverse diffusion
6	4–5 (Diffusion/Practice)	Scenarios, CRPS, calibration

Assessment idea. Ask participants to report MASE *and* CRPS for the seasonal-naive baseline, one neural/foundation model, and one generative model on a dataset of their choice, using a rolling-origin backtest — and to justify the model class they would ship.