STAT 9100 HW1

Mengyan Jing

2026-02-09

Instructions

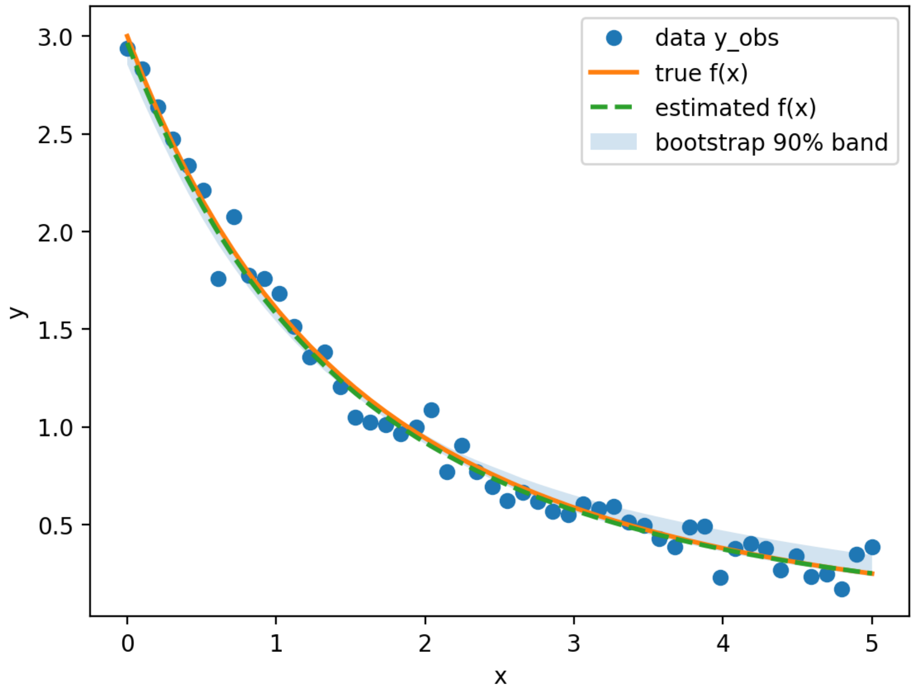
This homework explores neural estimation and conformal inference for a nonlinear regression model.

Problem 1

We trained a neural Bayes estimator to map the standardized observation vector $y \in \mathbb{R}^{50}$ to the parameter vector $\theta = (\beta_1, \gamma_1, \beta_2, \gamma_2)$ using simulated training pairs. To quantify uncertainty for the single observed curve, we used a residual bootstrap with 100 resamples.

Table 1: Neural Bayes estimate and bootstrap uncertainty (B=100).

Parameter	Estimate	BootMean	BootSD	CI2.5	CI97.5
beta1	1.522	1.530	0.076	1.375	1.669
gamma1	0.369	0.347	0.032	0.279	0.395
beta2	1.443	1.408	0.090	1.248	1.578
gamma2	1.010	1.058	0.089	0.908	1.244



Problem 2

Settings (Neural Likelihood)

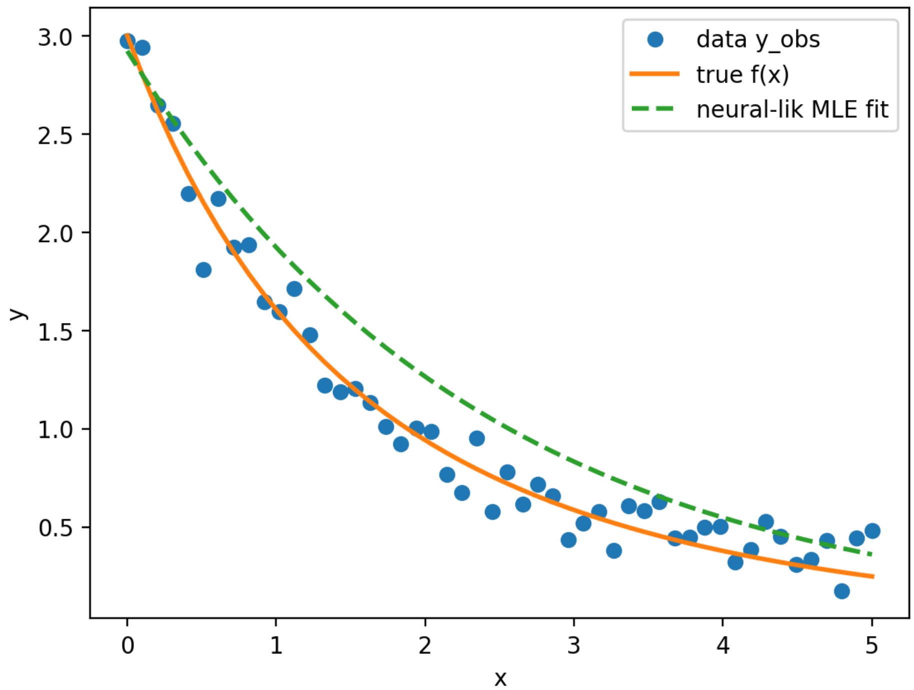
- Model: $y(x) = \beta_1 e^{-\gamma_1 x} + \beta_2 e^{-\gamma_2 x} + \epsilon$, $\epsilon \sim N(0, \sigma^2)$, $\sigma = 0.1$.
- Design points: x is a fixed grid of 50 points on $[0, 5]$.
- Prior (training simulations): $\beta_1, \beta_2 \sim \text{Unif}(0.5, 2.5)$, $\gamma_1, \gamma_2 \sim \text{Unif}(0.1, 1.6)$ with identifiability constraint $\gamma_1 < \gamma_2$.
- Training data: $N = 60{,}000$ simulated pairs.
- Classifier architecture: fully-connected network ($54 \to 128 \to 128 \to 1$) with ReLU.
- Training: Adam (lr=1e-3), batch size 512, epochs 8; response curves standardized using global mean/std from the simulated training set.
- MLE search: multi-start gradient ascent over a constrained parameterization (via sigmoid map to bounds and sorting to enforce $\gamma_1 < \gamma_2$); warm-started using Problem 1 estimate.

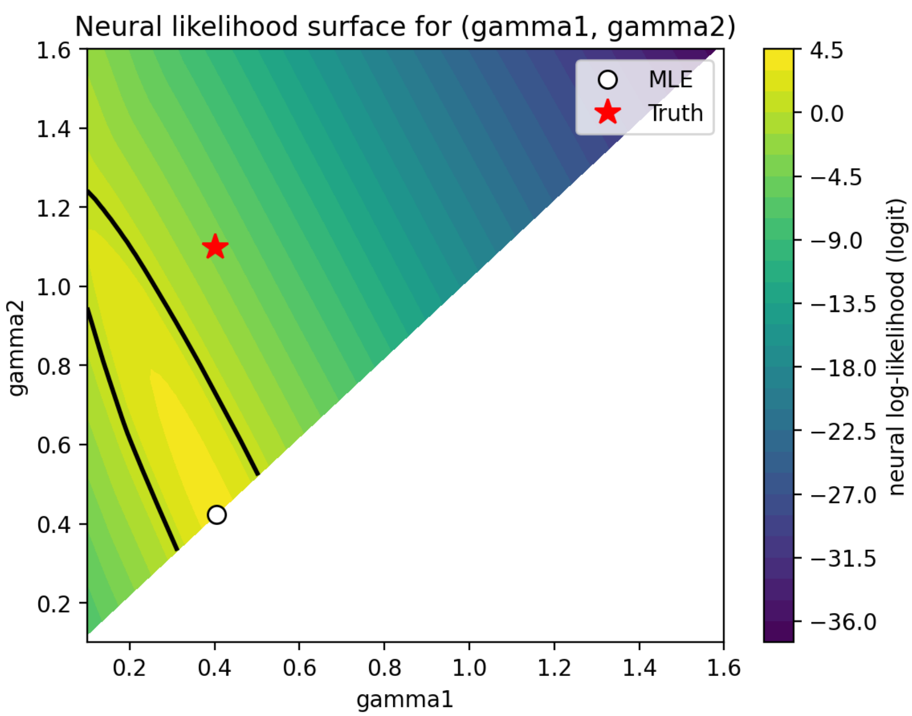
Table 2: Problem 2: Neural-likelihood approximate MLE and truth.

Parameter	Estimate	Truth
beta1	0.678	1.8
gamma1	0.405	0.4
beta2	2.245	1.2
gamma2	0.423	1.1

Table 3: Problem 2: Likelihood-ratio diagnostic at the truth.

best_logit	LR_at_truth	LR95_thresh_df2
3.854	3.263	5.99





Conclusion

We trained a neural likelihood classifier to distinguish *paired* samples $(y, \theta) \sim p(y \mid \theta)p(\theta)$ from *mismatched* samples obtained by permuting θ. The classifier logit evaluated at the observed curve y_{obs} provides an approximation to the log-likelihood surface $\ell(\theta; y_{\text{obs}})$ up to an additive constant. Maximizing this neural log-likelihood over a constrained parameterization (positivity and the identifiability constraint $\gamma_1 < \gamma_2$) yielded an approximate MLE

$$\hat{\theta}_{\text{MLE}} = (\hat{\beta}_1, \hat{\gamma}_1, \hat{\beta}_2, \hat{\gamma}_2) = (0.678,\ 0.405,\ 2.245,\ 0.423),$$

with best logit value 3.854. The fitted response captures the overall exponential decay pattern in the data, but deviates from the true response for larger x, suggesting that multiple parameter combinations can produce similar fitted curves in this two-exponential mixture model.

The two-dimensional neural likelihood surface over (γ_1, γ_2) exhibits an elongated ridge concentrated near $\gamma_1 \approx \gamma_2$, indicating weak identifiability of the two-exponential decomposition. This likelihood geometry explains why optimization can favor solutions with nearly equal decay rates and redistributed weights, even when the true decay rates are more separated. Uncertainty for (γ_1, γ_2) was quantified using a likelihood-ratio (LR) region based on the grid-evaluated surface. The LR statistic at the true parameters is

$$\text{LR}(\theta_{\text{true}}) = 3.263 < \chi_2^2(0.95) = 5.99,$$

so the true (γ_1, γ_2) lies inside the 95% LR region, consistent with the contour overlay in the likelihood surface plot.

Problem 3

Problem 4