MLE Sheet-2

May 2025

Exercise 3: Statistical Darts

Task 5

We model the dart thrower's aiming point $\mu \in R^2$ as a latent variable with:

- **Likelihood**: $p(x_i|\mu) = \mathcal{N}(\mu, \Sigma_{\text{true}})$ (Gaussian throws around μ)
- **Prior**: $p(\mu) = \mathcal{N}(\mu_0, \Sigma_0)$ (Gaussian belief about μ before seeing data)

The **posterior distribution** after observing N throws is Gaussian:

$$p(\mu|\{x_i\}) = \mathcal{N}(\mu_{\text{post}}, \Sigma_{\text{post}}), \tag{1}$$

where:

$$\Sigma_{\text{post}} = \left(\Sigma_0^{-1} + N\Sigma_{\text{true}}^{-1}\right)^{-1}, \tag{2}$$
$$\mu_{\text{post}} = \Sigma_{\text{post}} \left(\Sigma_0^{-1}\mu_0 + N\Sigma_{\text{true}}^{-1}\hat{\mu}_{\text{MLE}}\right). \tag{3}$$

Here, $\hat{\mu}_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the MLE estimate.

Role of the Prior

The prior influences the posterior through Σ_0 (prior covariance) and μ_0 (prior mean). We analyze three cases:

0.1 Strong (Concentrated) Prior

- **Prior**: $\Sigma_0 \to 0$ (e.g., $\Sigma_0 = \epsilon I$, $\epsilon \ll 1$)
- **Posterior**:

$$\Sigma_{\text{post}} \approx \Sigma_0$$
$$\mu_{\text{post}} \approx \mu_0$$

- **Interpretation**: The coach is *highly confident* that $\mu \approx \mu_0$. Data must be *overwhelming* to shift μ_{post} away from μ_0.

0.2 Weak (Diffuse) Prior

- **Prior**: $\Sigma_0 \to \infty$ (e.g., $\Sigma_0 = kI$, $k \gg 1$)
- **Posterior**:

$$\Sigma_{\text{post}} \approx \frac{\Sigma_{\text{true}}}{N}$$
$$\mu_{\text{post}} \approx \hat{\mu}_{\text{MLE}}$$

- **Interpretation**: The coach has *minimal preconceptions* about μ. Inference reduces to MLE-like behavior.

0.3 Non-Informative Prior

- **Prior**: $\Sigma_0^{-1} \to 0$ (improper prior, $p(\mu) \propto 1$)
- **Posterior**:

$\Sigma_{\text{post}} = \frac{\Sigma_{\text{true}}}{N}$

- **Interpretation**: Mathematically equivalent to MLE. No regularization; entirely data-driven.

Contrast with MLE

- **MLE**:

$$\hat{\mu}_{\text{MLE}} = \text{argmax}_\mu \, p(\{x_i\}|\mu)$$
$$E[\hat{\mu}_{\text{MLE}}] = \mu_{\text{true}}$$
$$\text{Cov}(\hat{\mu}_{\text{MLE}}) = \frac{\Sigma_{\text{true}}}{N}$$

  Limitation: No mechanism to incorporate prior knowledge.

- **Bayesian Inference**:
  - Estimate: μ_{post} balances prior (μ_0) and data ($\hat{\mu}_{\text{MLE}}$)
  - Uncertainty: Σ_{post} combines prior and data precision
  - Advantage: Naturally regularizes for small N

Implications

- **Strong Prior**:
  - Use for beginners expected to aim at bullseye
  - Conservative corrections; resists overreacting
- **Weak Prior**:
  - Use for experienced players with unique styles
  - Quickly adapts to observed throws
- **Non-Informative Prior**:
  - Fully data-driven coaching
  - Equivalent to MLE

Theoretical Takeaways

1. **Prior as Regularization**: Shrinks μ_{post} toward μ_0
2. **Uncertainty Quantification**: Σ_{post} reflects total uncertainty
3. **Asymptotic Behavior**: As $N \to \infty$, $\mu_{\text{post}} \to \hat{\mu}_{\text{MLE}}$
4. **Bias-Variance Tradeoff**:
   - Strong prior $\Rightarrow$ low variance, high bias
   - Weak prior $\Rightarrow$ high variance, low bias

Task 6

Conditions for Coincidence

From the posterior mean expression:

$$\mu_{\text{post}} = \Sigma_{\text{post}} \left(\Sigma_0^{-1} \mu_0 + N \Sigma_{\text{true}}^{-1} \hat{\mu}_{\text{MLE}} \right), \tag{4}$$

where $\Sigma_{\text{post}} = \left(\Sigma_0^{-1} + N \Sigma_{\text{true}}^{-1} \right)^{-1}$, the **MAP estimate** ($\mu_{\text{post}}$) coincides with the **MLE estimate** ($\hat{\mu}_{\text{MLE}}$) under two conditions:

1. Non-Informative Prior ($\Sigma_0^{-1} \to 0$)

- **Interpretation**: The prior carries no weight (infinite variance).
- **Effect**:
  - $\Sigma_0^{-1} \to 0$ cancels the prior term in μ_{post}
  - Posterior mean reduces to:
  $$\mu_{\text{post}} = \left(N \Sigma_{\text{true}}^{-1} \right)^{-1} \left(N \Sigma_{\text{true}}^{-1} \hat{\mu}_{\text{MLE}} \right) = \hat{\mu}_{\text{MLE}} \tag{5}$$
- **Intuition**: With no prior influence, inference relies entirely on data.

2. Infinite Data ($N \to \infty$)

- **Interpretation**: The data dominates the prior.
- **Effect**:
  - The term $N\Sigma_{\text{true}}^{-1}$ overwhelms Σ_0^{-1} in Σ_{post}:

$$\Sigma_{\text{post}} \approx \left(N\Sigma_{\text{true}}^{-1}\right)^{-1} = \frac{\Sigma_{\text{true}}}{N} \tag{6}$$

  - Posterior mean becomes:

$$\mu_{\text{post}} \approx \frac{\Sigma_{\text{true}}}{N}\left(N\Sigma_{\text{true}}^{-1}\hat{\mu}_{\text{MLE}}\right) = \hat{\mu}_{\text{MLE}} \tag{7}$$

- **Intuition**: Even a strong prior is overridden by sufficient data.

Key Observations

1. **Prior Irrelevance**:
   - In both cases, the prior term $\Sigma_0^{-1}\mu_0$ vanishes in μ_{post}
   - The MLE (pure data-driven estimate) emerges as the limiting case
2. **Uncertainty**:
   - Under these conditions, $\Sigma_{\text{post}} \to \frac{\Sigma_{\text{true}}}{N}$, matching the MLE's asymptotic covariance
3. **Practical Implications**:
   - For small N, MAP and MLE diverge unless the prior is non-informative
   - For large N, all reasonable estimators (MLE, MAP, posterior mean) converge

Task 7

Medical Diagnosis

- **Scenario**: Predicting patient disease risk from medical test results
- **Why Uncertainty Matters**:
  - False positives/negatives have serious consequences
  - Uncertainty estimates help clinicians:
    * Determine if additional tests are needed
    * Assess confidence in treatment decisions
    * Communicate risk probabilities to patients
- **Example**: A cancer prediction model with 60% confidence requires different follow-up than one with 95% confidence

Autonomous Vehicle Navigation

- **Scenario**: Real-time object detection and path planning
- **Why Uncertainty Matters**:
  - Safety-critical decisions require reliability estimates
  - High uncertainty triggers:
    * Reduced vehicle speed
    * Request for human intervention
    * Alternative sensor verification
- **Example**: A pedestrian detection with 80% uncertainty should cause more caution than one with 5% uncertainty