MAP & Bayesian Estimation

Parameter Estimation

Quiz

MLE vs MAP Objective

Let θ be a parameter and D the observed data. Which expression corresponds to the **MAP estimate** of θ?

A. $\hat{\theta}_{\mathrm{map}}$ = arg maxθ log p(D | θ)

B. $\hat{\theta}_{\mathrm{map}}$ = arg maxθ [log p(D | θ) + log p(θ)]

C. $\hat{\theta}_{\mathrm{map}}$ = arg maxθ log p(θ)

D. $\hat{\theta}_{\mathrm{map}}$ = arg maxθ log p(D)

Prior as Regularizer

The MAP objective is: $-\log p(D|\theta) + (-\log p(\theta))$. What is the main **effect** of including $-\log p(\theta)$ in the objective?

A. It always makes the model underfit.

B. It encourages parameter values favored by the prior, acting like a regularizer against overfitting.

C. It removes the need for a likelihood term.

D. It guarantees zero training error.

Posterior & Bayes' Rule

Which of the following correctly expresses the **posterior** over parameters in Bayesian inference?

A. $p(\theta|D) = p(D|\theta) \cdot p(\theta)$

B. $p(\theta|D) = p(D|\theta)p(\theta) / p(D)$

C. $p(\theta|D) = p(D)p(\theta) / p(D|\theta)$

D. $p(\theta|D) = p(\theta)$

Marginal Likelihood / Evidence

The **marginal likelihood** (also called evidence) p(D) in Bayesian inference is:

A. $p(D) = p(D \mid \hat{\theta}_{mle})$

B. $p(D) = p(D \mid \hat{\theta}_{map})$

C. $p(D) = \int p(D|\theta)\, p(\theta)\, d\theta$

D. $p(D) = \int p(\theta|D)\, p(\theta)\, d\theta$

Coin Flip: MLE vs MAP

You toss a coin 3 times and observe 3 heads, 0 tails. Let θ be the probability of heads. Likelihood: Bernoulli/Binomial in θ. Prior: θ ~ Beta(2,2). What is the **MAP** estimate of θ?

A. $\hat{\theta}_{map} = 1$

B. $\hat{\theta}_{map} = 0.5$

C. $\hat{\theta}_{map} = 0.75$

D. $\hat{\theta}_{map} = 0.8$

Pseudo-Observations Interpretation

In the Beta-Bernoulli setting, the prior $\theta \sim \text{Beta}(\alpha,\beta)$ can be interpreted as:

A. A belief that we have already seen $\alpha+\beta$ real coin tosses.

B. Pseudo-counts of $\alpha-1$ heads and $\beta-1$ tails (for $\alpha,\beta>1$), added to the data.

C. A constraint that $\alpha+\beta=1$.

D. Noise added to the observations.

Conjugate Priors

Which of the following **is NOT** a conjugate prior-likelihood pair?

A. Bernoulli likelihood + Beta prior

B. Poisson likelihood + Gamma prior

C. Multinomial likelihood + Dirichlet prior

D. Gaussian likelihood + Uniform prior over $\mathbb{R}$

Fully Bayesian Prediction vs Plug-in

For predicting a new observation x_new in a fully Bayesian approach, which expression is correct?

A. $p(x\_new \mid D) = p(x\_new \mid \hat{\theta}_{mle})$

B. $p(x\_new \mid D) = p(x\_new \mid \hat{\theta}_{map})$

C. $p(x\_new \mid D) = \int p(x\_new \mid \theta)\, p(\theta \mid D)\, d\theta$

D. $p(x\_new \mid D) = p(\theta \mid D)$

Online Nature of Bayesian Inference

Bayesian inference has an "online" nature. What does that mean in practice?

A. The posterior does not depend on the order of data arrival.

B. We can treat the current posterior as the new prior when additional data arrive, and update by the same Bayes' rule.

C. Bayesian methods require data to arrive in real time.

D. The posterior becomes exactly uniform after many observations.

Effect of More Data on the Posterior

As the number of observations N grows very large, which of the following is typically true (under mild regularity conditions)?

A. The posterior over θ becomes more spread out.

B. The prior completely dominates, making the data almost irrelevant.

C. The posterior becomes increasingly concentrated near the MLE solution.

D. The posterior always becomes uniform.