Bayesian Linear Regression: Complete Hand Calculation Example

Problem Setup

We'll work through a simple linear regression problem by hand:

- Model: $y = a + bx + \varepsilon$
- Prior: $w \sim N(0, (1/\alpha)I)$ where $w = [a, b]^T$
- Noise: $\varepsilon \sim N(0, 1/\beta)$

Data

We have just 2 data points to keep calculations manageable:

- Point 1: $(x_1, y_1) = (0, 1)$
- Point 2: $(x_2, y_2) = (1, 3)$

Hyperparameters

- $\alpha = 1$ (prior precision)
- $\beta = 4$ (noise precision, so $\sigma^2 = 0.25$)

Step 1: Set Up Design Matrix

The design matrix Φ includes a column of ones (for intercept) and x values:

```
Φ = [1  x₁]   = [1  0]
    [1  x₂]     [1  1]
```

Step 2: Set Up Prior

Prior distribution: $w \sim N(m_0, S_0)$

- Mean: $m_0 = [0, 0]^T$
- Covariance: $S_0 = (1/\alpha)I = (1/1)I = I$

```
S₀ = [1  0]
     [0  1]
```

Inverse: $S_0^{-1} = \alpha I = I$

Step 3: Calculate $\Phi^T \Phi$

```
Φᵀ = [1  1]
     [0  1]

Φᵀ Φ = [1  1] × [1  0] = [1×1+1×1  1×0+1×1] = [2  1]
       [0  1]   [1  1]   [0×1+1×1  0×0+1×1]   [1  1]
```

Step 4: Calculate Posterior Covariance

Formula: $\boxed{S = (S_0^{-1} + \beta\Phi^T\Phi)^{-1}}$

First, calculate $\boxed{S_0^{-1} + \beta\Phi^T\Phi}$:

```
S₀⁻¹ + βΦᵀΦ = [1  0] + 4×[2  1] = [1  0] + [8  4] = [9  4]
              [0  1]     [1  1]   [0  1]   [4  4]   [4  5]
```

Now invert this 2×2 matrix:

```
S⁻¹ = [9  4]
      [4  5]

det(S⁻¹) = 9×5 - 4×4 = 45 - 16 = 29

S = (1/29) × [5  -4] = [5/29   -4/29]
             [-4  9]   [-4/29   9/29]
```

Numerically:

```
S = [0.172  -0.138]
    [-0.138  0.310]
```

Step 5: Calculate $\Phi^T y$

```
y = [1]
    [3]

Φᵀ y = [1  1] × [1] = [1×1 + 1×3] = [4]
       [0  1]   [3]   [0×1 + 1×3]   [3]
```

Step 6: Calculate Posterior Mean

Formula: $m = S(S_0^{-1}m_0 + \beta\Phi^T y)$

Since $m_0 = 0$, we have:

```
m = S(βΦᵀy) = S × 4 × [4] = S × [16]
                      [3]       [12]
```

```
m = [5/29   -4/29] × [16] = [(5×16 - 4×12)/29]  = [(80-48)/29]  = [32/29]
    [-4/29   9/29]   [12]   [(-4×16 + 9×12)/29]   [(-64+108)/29]  [44/29]
```

Numerically:

```
m = [1.103]
    [1.517]
```

Interpretation

The posterior mean gives us:

- Intercept (a): 1.103
- Slope (b): 1.517

So our estimated line is: $y = 1.103 + 1.517x$

Step 7: Make a Prediction

Let's predict at x* = 0.5:

```
ϕ* = [1, 0.5]
```

Mean Prediction

```
μ* = ϕ*ᵀm = [1, 0.5] × [1.103] = 1.103 + 0.5×1.517 = 1.862
                       [1.517]
```

Prediction Variance

Function variance:

```
σ²ₓ = φ*ᵀSφ*
```

First calculate `Sφ*`:

```
Sφ* = [0.172  -0.138] × [1  ] = [0.172 - 0.069]  = [0.103]
      [-0.138  0.310]   [0.5]   [-0.138 + 0.155]   [0.017]
```

Then:

```
σ²ₓ = [1, 0.5] × [0.103] = 0.103 + 0.5×0.017 = 0.103 + 0.0085 = 0.1115
                 [0.017]
```

Predictive variance (including noise):

```
σ²ᵧ = σ²ₓ + 1/β = 0.1115 + 1/4 = 0.1115 + 0.25 = 0.3615
```

Standard deviation: `σᵧ = √0.3615 = 0.601`

Step 8: 95% Credible Intervals

For the parameters:

```
a: 1.103 ± 1.96×√0.172 = 1.103 ± 1.96×0.415 = 1.103 ± 0.813
   → [0.290, 1.916]

b: 1.517 ± 1.96×√0.310 = 1.517 ± 1.96×0.557 = 1.517 ± 1.092
   → [0.425, 2.609]
```

For the prediction at $x^* = 0.5$:

```
y*: 1.862 ± 1.96×0.601 = 1.862 ± 1.178
    → [0.684, 3.040]
```

Summary

From just 2 data points:

- Estimated line: `y = 1.103 + 1.517x`
- Prediction at x=0.5: 1.862 ± 1.178 (95% CI)

- Intercept uncertainty: ±0.813
- Slope uncertainty: ±1.092

The calculations show how:

1. Prior information (S_0) combines with data information ($\beta\Phi^T\Phi$)
2. More data points would make S^{-1} larger, reducing posterior variance
3. Predictions include both parameter uncertainty and noise

Verification Check

Let's verify our line passes near the data:

- At x=0: y = 1.103 (actual: 1) ✓
- At x=1: y = 1.103 + 1.517 = 2.620 (actual: 3) ✓

The slight differences are expected due to noise and prior influence.