3. Bayesian Estimation and Linear Regression

We will now explore what linear regression looks like from a particular Bayesian Framework. The answers that you find here may be surprising to you, hopefully in a pleasant way.

Suppose that:

- $Y_1, \ldots, Y_n$ are independent given the pair (β_0, β_1)
- $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where each ϵ_i are i.i.d. $\mathcal{N}(0, 1/\tau)$ (which has variance $1/\tau$)
- the $X_i \in \mathbb{R}$ are deterministic.

We will think of β_0, β_1 and τ as being random variables.

Suppose that we place an improper prior on (β_0, β_1, τ):

$$\pi(\beta_0, \beta_1, \tau) = 1/\tau.$$

Since this expression on the right hand side does not depend on the β's, we may take the conditional distribution $\pi(\beta_0, \beta_1 | \tau)$ to be the "uniform" improper prior: $\pi(\beta_0, \beta_1 | \tau) = 1$.

Answer the following problems given these assumptions. As a reminder, we let $\mathbb{X}$ be the design matrix, where the ith row is the row vector $(1, X_i)$, and let $\mathbf{Y}$ be the column vector $\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$.

(a) The Bayesian setup: The posterior distribution

2.0/2 points (graded)

Observe that if β_0, β_1 and τ are given, then each Y_i is a gaussian: $Y_i | (\beta_0, \beta_1, \tau) \sim \mathcal{N}(\beta_0 + \beta_1 X_i, 1/\tau)$.

Therefore, the likelihood function of the vector $(Y_1, \ldots, Y_n)$ given (β_0, β_1, τ) is of the form

$$\left(\frac{1}{\sqrt{2\pi/\tau}}\right)^n \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i)^2\right)$$

It turns out that the distribution of (β_0, β_1) given τ and $Y_1, \ldots, Y_n$ is a 2-dimensional Gaussian. In terms of $\mathbb{X}$, $\mathbf{Y}$ and τ, what is its mean and covariance matrix?

Hint: look ahead and see what part (b) is asking. What answer do you hope would come out, at least for one of these two things?

(Type **X** for $\mathbb{X}$, **trans(X)** for the transpose $\mathbb{X}^T$, and **X^(-1)** for the inverse $\mathbb{X}^{-1}$ of a matrix $\mathbb{X}$.)

Mean: (trans(X)*X)^(-1)*trans(X)*Y ✔ **Answer:** (trans(X)*X)^(-1)*trans(X)*Y

Covariance: (trans(X)*X)^-1/tau ✔ **Answer:** (trans(X)*X)^(-1)/tau

STANDARD NOTATION

Solution:

Apply Baye's Rule:

$$\begin{aligned}
\pi(\beta_0, \beta_1 | \tau, \mathbf{Y}) &= \frac{\pi(\mathbf{Y}|\beta_0, \beta_1, \tau) \cdot \pi(\beta_0, \beta_1|\tau)}{\pi(\mathbf{Y}|\tau)} \\
&\propto \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right) \quad \text{prior} \times \text{likelihood function} \\
&= \exp\left(-\frac{\tau}{2}\|\mathbf{Y} - \mathbb{X}\beta\|_2^2\right) \\
&= \exp\left(-\frac{\tau}{2}(\mathbf{Y} - \mathbb{X}\beta)^T(\mathbf{Y} - \mathbb{X}\beta)\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\beta^T(\tau\mathbb{X}^T\mathbb{X})\beta - 2\tau\mathbf{Y}^T\mathbb{X}\beta\right]\right)
\end{aligned}$$

From here, it can be tricky to figure out what the right μ and Σ ought to be, but there are a few ways to figure it out.

One method is to notice that the multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ has density $f(x; \mu, \Sigma)$ proportional to $\exp\left(-(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$. In our setting, β plays the role of x, so there ought to be a $\beta^T\Sigma^{-1}\beta$ term somewhere. This allows us to take $\Sigma^{-1} = \tau\mathbb{X}^T\mathbb{X}$, and the only possible choice that makes the calculation work is $\mu = \frac{\Sigma\mathbb{X}^T\mathbf{Y}}{\tau} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}$.

Notice that μ here $\boxed{\text{coincides with}}$ the least squares estimator. In fact, this was the provided hint: looking forward, part (b) asks to find the Bayes estimator of (β_0, β_1), which is exactly μ (since the distribution is a Gaussian). This may feel like a bit of a cheat: *it would be very nice if the LSE $\hat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}$ were actually the answer*, so one could simply guess this expression for μ and derive Σ accordingly. This is a second way that one can arrive at the desired answer.

Submit You have used 2 of 4 attempts

Answers are displayed within the problem

(b)

1.0/1 point (graded)

What is the Bayes estimator $\widehat{(\beta_0, \beta_1)}^{\text{Bayes}}$ for (β_0, β_1)?

Hint: Use your answer from part (a). However, as hinted: the answer here is guessable, even if you didn't solve the previous part.

(Answer in terms of $\mathbb{X}$, $\mathbf{Y}$ and τ.)

(Type **X** for $\mathbb{X}$, **trans(X)** for the transpose $\mathbb{X}^T$ of a matrix $\mathbb{X}$, and **X^(-1)** for the inverse $\mathbb{X}^{-1}$ of a matrix $\mathbb{X}$.)

$\widehat{(\beta_0, \beta_1)}^{\text{Bayes}} =$ (trans(X)*X)^(-1)*trans(X)*Y

Answer: (trans(X)*X)^(-1)*trans(X)*Y

STANDARD NOTATION

Solution:

The Bayes estimator is the mean itself, since the likelihood function of a Gaussian is maximized at the mean.

Notice that this answer agrees with the Least Squares Estimator. In some sense, this is not a coincidence – the improper prior was designed to make this work out. *Think, from a high-level perspective: why would a uniform distribution for $(\beta_0, \beta_1)|\tau$ translate into these answers agreeing?*

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(c)

1/1 point (graded)
Given our improper prior, we ought to take the posterior distribution of $\tau | (\beta_0, \beta_1)$ to also be $\pi(\tau|\beta_0,\beta_1) = \frac{1}{\tau}$, for each realization of τ.

What type of distribution is the posterior distribution of τ given the **triple** $(\beta_0, \beta_1, \mathbf{Y})$?

- Gaussian
- Chi-Squared
- Gamma ✔
- Uniform
- Exponential
- Other (not listed above)

Solution:

To make things look nicer, let S^2 denote the quantity

$$S^2 := \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2.$$

This makes it so that the posterior distribution of τ looks like, via Baye's Rule,

$$\pi(\tau|\beta_0,\beta_1,\mathbf{Y}) \;=\; \frac{\pi(\mathbf{Y}|\tau,\beta_0,\beta_1)\,\pi(\tau|\beta_0,\beta_1)}{\pi(Y|\beta_0,\beta_1)}$$

The denominator does not depend on τ, so we absorb it into a constant of proportionality:

$$\propto \tau^{n/2}\exp\left(-\tfrac{\tau}{2}S^2\right)\cdot\frac{1}{\tau}$$

which describes a Gamma distribution, $\mathbf{Gamma}\left(\frac{n}{2}, \frac{S^2}{2}\right)$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 6 Linear Regression:Homework 10 Linear regression / 3. Bayesian Estimation and Linear Regression