
Models

Justin Salmon¹ and George Lancaster²

¹wr18313

²qv18258

October 17, 2018

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce maximus nisi ligula. Morbi laoreet ex ligula, vitae lobortis purus mattis vel. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec ac metus ut turpis mollis placerat et nec enim. Duis tristique nibh maximus faucibus facilisis. Praesent in consequat leo. Maecenas condimentum ex rhoncus, elementum diam vel, malesuada ante.

1 The Prior

1.1 Theory

Question 1.1

A Gaussian likelihood encodes the inherent noise present in most real-world data. In other words, most probabilistic processes in nature tend to be noisy, and that noise tends to follow a Gaussian distribution. Hence this is generally a good place to start.

Noise in input data - assume noise is gaussian -> implies gaussian likelihood

Question 1.2

Choosing a spherical covariance matrix essentially means that we are assuming that the input noise is going to be equal in all directions, i.e. the distribution is equally likely to deviate from the mean in all directions. Again, this is a good place to start. Choosing a non-spherical covariance would imply that we know something in advance about the relationship between the input and output, which is not true in this case.

Question 2

TODO The covariance matrix would not be in terms of the identity matrix (for some reason). We would

have non-zero values in the offset diagonals which correspond to the correlations between different things.

1.2 Linear Regression

Question 3

$$p(\mathbf{Y}|\mathbf{X}, \mathbf{W}, \beta) = \prod_{i=0}^N \mathcal{N}(y_i | \mathbf{W}^T \phi(x_i), \beta^{-1}) \quad (1)$$

Question 4

A distribution is conjugate to another if they both take the same algebraic form, i.e. are in the same probability distribution family. For example, Gaussians are conjugate to each other, and the conjugate to a Bernoulli distribution is a Beta distribution. Conjugates are used as a convenience used to avoid calculating the denominator in Bayes' rule (the evidence) which can often be a tricky integral. If the prior and likelihood are conjugate, then their product will be proportional to the posterior.

Question 5

Euclidean distance

Question 6

Derive posterior mean and covariance, start at conjugacy. Watch video <https://www.youtube.com/watch?v=nrd4AnDLR3U>

1.3 Non-parametric Regression

2 Posterior

$$f \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I}) \quad (2)$$

$$\beta \sim \Gamma(a) \quad (3)$$

3 Evidence