From Eriks request I have tried to summarise some help to the questions that I most commonly get.

0.1 Q6. Derive the posterior

The trick here is first to look at one dimension of the problem and assume that each point are independent, so if y is one dimensional we will have something like this,

$$p(\mathbf{y}|\mathbf{W},\mathbf{X}) = \prod_{i=1}^{N} p(y_i|\mathbf{W},\mathbf{X}) \propto exp((\mathbf{y}-\mathbf{W}\mathbf{x})^{\mathrm{T}}\Sigma^{-1}(\mathbf{y}-\mathbf{W}\mathbf{x})) \tag{1}$$

Then you want to add additional dimensions to $\mathbf{y}$ this means that you will now have a matrix in the exponential. Think about the elements of this matrix and you will realise that it is the sum of the diagonal that really makes sense, i.e. you want to take the trace of the whole thing.

0.2 Q8.

So, I might have given a bit too much information here which confused more than it helped. I just simply want you to explain why this is a sensible prior. Think about it this way, we need to somehow parametrise a prior, different priors give different possibilities due to their parametrisation. Is this one sensible/good in that way?

0.3 Q11.

You want to show just the images that are on the slides in my first lecture.

0.4 Q14.

You want to show similar plots to the ones that I showed in Lecture 2 where I sampled from the posterior.

0.5 Q17.

Have a look in the book Eq. 12.37 and Eq. 12.38 and the text surrounding it. So use the laws of Expectations and you should be able to do this fairly straight forward.

0.6 Q19.

This seems to be the biggest hurdle so far. So first take the derivation that you did in Q17. that you should now take the negative logarithm of which will give us the objective function. The tricky bit is now computing the derivatives. So what we need to do here is use [1].

The first thing to note is that we want to have a derivative for each element in our weight matrix. To do this the easiest way is to start writing the derivatives in terms of each element of the matrix. So, create yourself a nested for-loop that cycles through each element and we now need to compute this expression,

$$\frac{\delta \mathcal{L}}{\delta \mathbf{W}_{ij}}. \tag{2}$$

This will lead you to two terms, one containing a determinant and one other. The derminant term we can just look up in [1] and the problematic term is the other. To do this we can use the chain rule several times until we get to a derivative of an inverse to see what that one we use equation 40 in [1]. Now we have an expression which is quadratic i $\mathbf{W}$ so again you just apply the chain rule on this and you will get two terms. In the end you will have a simple derivative,

$$\frac{\delta \mathbf{W}}{\delta \mathbf{W}_{ij}} = \mathbf{J}_{ij} \tag{3}$$

where J_{ij} is a matrix of the same size as $\mathbf{W}$ with all zeros except for the element with ij which is 1.

Using this information you should be able to compute the derivative.

[1]K. Petersen and M. Pedersen, The matrix cookbook, Technical University of Denmark, 2006.