

Week 10 exercises

This is the seventh page of *assessed* questions, as described in the background notes. These questions form 70% of your mark for Week 10. The introductory questions in the notes and the Week 10 discussion group task form the remaining 30% of your mark for Week 10.

Unlike the questions in the notes, you'll not immediately see any example answers on this page. However, you can edit and resubmit your answers as many times as you like until the deadline (Friday 27 November 4pm 6pm UK time, UTC). This is a *hard deadline*: This course does not permit extensions and any work submitted after the deadline will receive a mark of zero. See the late work policy.

Queries: Please don't discuss/query the assessed questions on hypothesis until after the deadline. If you think there is a mistake in a question this week, please email Iain.

Please only answer what's asked. Markers will reward succinct to-the-point answers. You can put any other observations in the "Add any extra notes" button (but this is for your record, or to point out things that seemed strange, not to get extra credit). Some questions ask for discussion, and so are open-ended, and probably have no perfect answer. For these, stay within the stated word limits, and limit the amount of time you spend on them (they are a small part of your final mark).

Feedback: We'll return feedback on your submission via email by Friday 4 December.

Good Scholarly Practice: Please remember the University requirements for all assessed work for credit. Furthermore, you are required to take reasonable measures to protect your assessed work from unauthorised access. For example, if you put any such work on a public repository then you must set access permissions appropriately (permitting access only to yourself). You may not publish your solutions after the deadline either.

1 Linear and non-linear autoencoders

We centre our data so it has zero mean and fit a linear autoencoder with no bias parameters. The autoencoder is a D -dimensional vector-valued function \mathbf{f} , computed from D -dimensional inputs \mathbf{x} , using an intermediate K -dimensional "hidden" vector \mathbf{h} :

$$\begin{aligned}\mathbf{h} &= W^{(1)}\mathbf{x} \\ \mathbf{f} &= W^{(2)}\mathbf{h}.\end{aligned}$$

Assume we want to find a setting of the parameters that minimizes the square error $\|\mathbf{f} - \mathbf{x}\|^2 = (\mathbf{f} - \mathbf{x})^\top (\mathbf{f} - \mathbf{x})$, averaged (or summed) over training examples.

- a) The PCA solution sets $W^{(1)} = V^\top$ and $W^{(2)} = V$, where the columns of V contain eigenvectors of the covariance of the inputs. We only really need to fit one matrix to minimize square error. Tying the weight matrices together: $W^{(1)} = U^\top$ and $W^{(2)} = U$, we can fit one matrix U .

Give a series of equations that show how to compute \bar{U} by reverse-mode differentiation (or "backprop"). Here $\bar{U}_{ij} = \frac{\partial c}{\partial U_{ij}}$, where $c = (\mathbf{f} - \mathbf{x})^\top (\mathbf{f} - \mathbf{x})$ is the cost for a single input.

List the equations in the order that we would compute them. Assume that we have already performed a forward pass of the network, and have computed and stored, \mathbf{h} and \mathbf{f} for the current input \mathbf{x} . [15 marks]

[The website version of this note has a question here.]

- b) For the following questions i-iii), assume some 2-dimensional points lie along the one-dimensional circumference of a semi-circle. You could create such a dataset, by drawing one of the features from a uniform distribution between -1 and $+1$, and setting the other feature based on that:

$$x_1^{(n)} \sim \text{Uniform}[-1, 1]$$
$$x_2^{(n)} = \sqrt{1 - (x_1^{(n)})^2}.$$

- i) Explain why these points can't be perfectly reconstructed when passed through the linear autoencoder defined above with $K=1$. [10 marks]

[The website version of this note has a question here.]

- ii) Explain whether the points could be perfectly reconstructed with $K=1$ by some non-linear decoder: $\mathbf{f} = \mathbf{g}(h)$. Where \mathbf{g} could be an arbitrary function, perhaps represented by multiple neural network layers. Assume the encoder is still linear: $h = W^{(1)}\mathbf{x}$. [10 marks]

[The website version of this note has a question here.]

- iii) Explain whether the points could be perfectly reconstructed with $K=1$ when using some non-linear encoder: $h = g(\mathbf{x})$. Where g could again be an arbitrary function, perhaps represented by multiple neural network layers. Assume the decoder is still linear: $\mathbf{f} = W^{(2)}h$. [10 marks]

[The website version of this note has a question here.]

- c) Discuss advantages and disadvantages of the four different high-level architectures for an autoencoder:

- i) linear encoder, linear decoder
- ii) linear encoder, non-linear decoder
- iii) non-linear encoder, linear decoder
- iv) non-linear encoder, non-linear decoder

[25 marks]

[The website version of this note has a question here.]