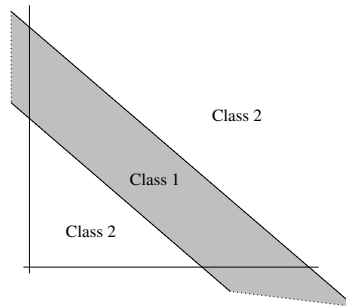


10 Machine Learning and Bayesian Inference (sbh11)

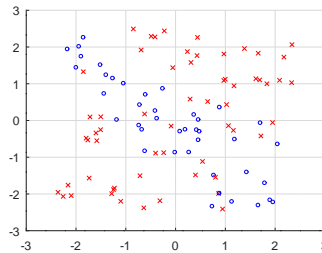
The standard linear classifier for two-class problems models the generation of noisy data as

$$\Pr(\text{Class 1}|\mathbf{w}, b; \mathbf{x}, \theta) = \sigma_{\theta}(\mathbf{w}^T \mathbf{x} + b) \quad (1)$$

where $\sigma_{\theta}(x) = 1/(1 + \exp(-\theta x))$. You are presented with a problem where the data appears not quite linear, in the sense that Class 1 forms a band, with Class 2 on either side:



The data is still noisy however, and looks like this:



We refer to this as the *banded* data.

- (a) Explain the purpose of each element of (1). Explain why each parameter is needed, and how each relates to the modelling of noisy, linear data. [6 marks]
- (b) Explain how the model in (1) can be modified to model the banded data, by leaving the linear part of the model unchanged, but modifying the treatment of the noise. [3 marks]
- (c) Explain how the model in (1) can be modified to model the banded data, by making a minimal modification to the linear part of the model, while leaving the treatment of the noise unchanged. [5 marks]
- (d) Assuming that you have training algorithms suitable for both of the models proposed for the banded data, briefly explain how, in practice, you might choose which is the better one to use, and how you might choose any relevant hyperparameters. [6 marks]