HW # 2

1. [Bayesian curve fitting and model selection]

Generate $N = 200$ datapoints from a mixture of two 2D gaussians:

$$G(\vec{x}) = p_1 \mathcal{N}(\vec{x} \mid \vec{\mu}_1, \Sigma) + p_2 \mathcal{N}(\vec{x} \mid \vec{\mu}_2, \Sigma),$$

where $\begin{cases} \vec{\mu}_1 = (0,1) \\ \vec{\mu}_2 = (1,0) \end{cases}$ $\quad \Sigma = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\sigma = 2$

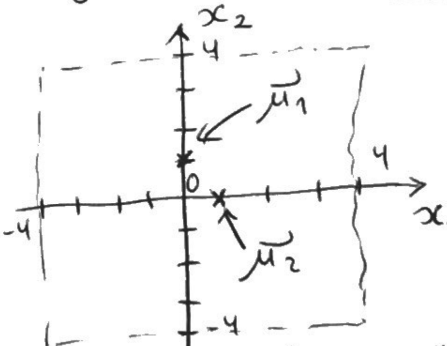
$p_1 = 0.3, \quad p_2 = 0.7$

Add noise: $\mathcal{N}(0, \beta^{-1})$ and assume $\beta = 100$ is known.

Use a gaussian basis with $2^2, 3^2, 4^2, 5^2, 6^2, 7^2$ 2D gaussians of the form $\mathcal{N}(\vec{x} \mid \vec{m}_j, \mathbb{I})$, ($\mathbb{I}$ — unit matrix)

where $\vec{m}_j$ uniformly cover the square area shown:

$\begin{cases} -4 \le x_1 \le 4 \\ -4 \le x_2 \le 4 \end{cases}$



[Choose $\vec{m}_j$ arbitrarily but "uniformly" within this area]

Using $\{t_n, \vec{x}_n\}_{n=1}^{200}$ as the input dataset, find the predictive distribution $p(t \mid \vec{x})$ under the evidence approximation (that is, find $\hat{\alpha}$ & use known β).

Plot the mean and the variance of the predictive distribution as heat maps in the [-4, 4] square shown above. Repeat for each of the 6 basis sets.

Using $\hat{\alpha}$ inferred for each basis set, plot the model evidence vs. the total number of gaussians in the basis for each of the 6 basis sets. Which model would you choose?

$$\log p(\vec{t} \mid \hat{\alpha}, \beta)$$

$\vec{t} = (t_1 \dots t_N)$

(2.) Bishop 4.9

(3.) Bishop 4.10