Example Sheet 4

11 EMCEE

(a) As discussed in the lecture, the EMCEE stretch move proposes a new candidate by sampling from $P(x) \propto \frac{1}{\sqrt{x}}$ for $x \in [0.5, 2]$. What is the probability of proposing a contraction ($x < 1$)? What is the probability of proposing an expansion ($x > 1$)?

(b) How do the probabilities of proposing contraction/expansion influence the ability of EMCEE to converge from an initial guess to the posterior maximum?

(c) Make an educated guess how the number of EMCEE walkers should roughly scale with number of fit parameters, i.e., with dimensionality of the parameter space that the EMCEE should explore.

12 Bias-variance decomposition

(a) Given that $\chi^2 \propto \langle (Y - f(X))^2 \rangle$, show that one can decompose $\chi^2 \propto \langle (Y - Y_{\text{true}})^2 \rangle + (Y_{\text{true}} - f(X))^2$.

(b) Explain why the first term is the variance and the second term the squared bias.

(c) Discuss why χ^2-minimisation can suffer from overfitting (provided the model is sufficiently flexible).

13 Bayesian evidence – Peaks of likelihood and prior

Consider a linear model with conjugate prior given by

$$\log P(\vec{\theta}) = -\frac{1}{2}(\vec{\theta} - \vec{\theta}_0)^2$$

that is obviously centred at $\vec{\theta}_0$ and has covariance matrix of $\Sigma_0 = I$. The likelihood of the linear model is a multivariate Gaussian whose maximum is located at

$$\vec{\theta}_{\text{MLE}} = \left(D^T \cdot \Sigma^{-1} \cdot D\right)^{-1} \cdot D^T \cdot \Sigma^{-1} \cdot \vec{y}$$

and whose covariance matrix can be found from a Fisher analysis:

$$\Sigma_\theta = \left(D^T \cdot \Sigma^{-1} \cdot D\right)^{-1}$$

If the prior peak is far away from the likelihood peak, evidence estimation via Monte-Carlo sampling is extremely inefficient.

(a) Assume that we only have a single parameter $\vec{\theta} = \theta$ and that $\theta_{\text{MLE}} = 0$ and $\Sigma_\theta = \sigma_\theta^2 = 1$. What is the chance of randomly drawing a θ from the 1D Gaussian prior that falls into $\theta_{\text{MLE}} \pm \sigma_\theta$ (near the likelihood peak) if the prior is centred at $\theta_0 = 1$, $\theta_0 = 3$ and $\theta_0 = 10$? (Write down the integral ansatz and evaluate the integral, e.g., using Wolfram Alpha.)

(b) Bonus question: Let us assume we have two parameters $\vec{\theta} = (\theta_1, \theta_2)^T$ and $\vec{\theta}_{\text{MLE}} = \vec{0}$ and $\Sigma_\theta = I$. What is the chance of drawing a $\vec{\theta}$ from the 2D Gaussian prior that falls into $|\vec{\theta} - \vec{\theta}_{\text{MLE}}|^2 \leq 1$ (near the likelihood peak) if the prior is centred at $\vec{\theta}_0 = (10, 0)^T$? (Write down the integral ansatz and evaluate the integral, e.g., using Wolfram Alpha.) Compare to the 1D case.