Hw1

1. Coin Tossing

   We have two coins, one fair and one biased. The probability of bringing heads with the biased coin is $\frac{1}{20}$. We close our eyes and choose one of the two coins and we toss it twice. Each coin has probability $\frac{1}{2}$ of being chosen. Compute the probability of:

   1. bringing heads in the first toss.
   2. having chosen the fair coin given that both tosses were heads.

Hw2

2. Consider a linear basis function regression model for a multivariate target variable t having a Gaussian distribution of the form

$$p(\mathbf{t}|\mathbf{W}, \Sigma) = \mathcal{N}(\mathbf{t}|\mathbf{y}(\mathbf{x}, \mathbf{W}), \Sigma),$$

where

$$\mathbf{y}(\mathbf{x}, \mathbf{W}) = \mathbf{W}^T \boldsymbol{\phi}(\mathbf{x}),$$

together with a training data set comprising input basis vectors $\phi(x_n)$ and corresponding target vectors $\mathbf{t}_n$, with $n = 1, \cdots, N$. Show that the maximum likelihood solution $\mathbf{W}_{ML}$ for the parameter matrix $\mathbf{W}$ has the property that each column is given by an expression of the form

$$\mathbf{W}_{ML} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t},$$

which was the solution for an isotropic noise distribution. Note that this is independent of the covariance matrix Σ. Show that the maximum likelihood solution for Σ is given by

$$\Sigma = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{t}_n - \mathbf{W}_{ML}^T \phi(x_n))(\mathbf{t}_n - \mathbf{W}_{ML}^T \phi(x_n))^T.$$

Hw3

3. Show that the logistic sigmoid function

$$\sigma(a) = \frac{1}{1 + \exp(-a)}, \tag{1}$$

satisfies the property $\sigma(-a) = 1 - \sigma(a)$ and that its inverse is given by $\sigma^{-1}(y) = \ln\{y/(1-y)\}$

Hw4

4. By considering the determinant of a 2×2 Gram matrix, show that a positive-definite kernel function $k(x, x')$ satisfies the Cauchy-Schwartz inequality

$$k(x_1, x_2)^2 \leq k(x_1, x_1)k(x_2, x_2). \tag{2}$$

5. Using the Newton-Raphson formula (**??**), stated as follows

$$\mathbf{w}^{(\text{new})} = \mathbf{w}^{(\text{old})} - \mathbf{H}^{-1}\nabla E(\mathbf{w}), \tag{3}$$

to derive the following iterative update formula:

$$a_N^{\text{new}} = \mathbf{C}_N(\mathbf{I} + \mathbf{W}_N\mathbf{C}_N)^{-1}\{\mathbf{t}_N - \sigma_N + \mathbf{W}_N a_N\} \tag{4}$$

for finding the model a_N^* of the posterior distribution in the Gaussian process classification model.