Machine Learning and Bayesian Inference

Problem Sheet I

Sean B. Holden

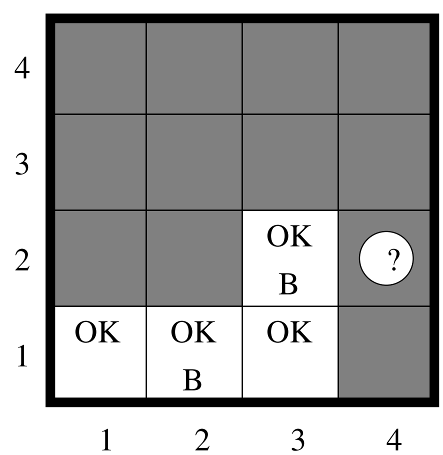
1 Basic probability: warm-up question

1. This question revisits the Wumpus World, but now our valiant hero, having learned the importance of probability by attending *Machine Learning and Bayesian Inference*, will use probabilistic reasoning instead of the situation calculus.

   Through careful consideration of the available knowledge on Wumpus caves, the explorer has established that each square contains a pit with probability 0.3, and pits are independent of one-another. Let $\texttt{Pit}_{i,j}$ be a Boolean random variable (RV) having values in $\{\top, \bot\}$ and denoting the presence of a pit at row i, column j. So for all (i, j)

   $$\Pr\left(\texttt{Pit}_{i,j} = \top\right) = 0.3$$
   $$\Pr\left(\texttt{Pit}_{i,j} = \bot\right) = 0.7.$$

   In addition, after some careful exploration of the current cave, the explorer has discovered the following:



   $$\texttt{Pit}_{1,1} = \bot$$
   $$\texttt{Pit}_{1,2} = \bot$$
   $$\texttt{Pit}_{1,3} = \bot$$
   $$\texttt{Pit}_{2,3} = \bot$$

   B denotes squares where a breeze is perceived. Let $\texttt{Breeze}_{i,j}$ be a Boolean RV denoting the presence of a breeze at (i, j)

   $$\texttt{Breeze}_{1,2} = \texttt{Breeze}_{2,3} = \top$$
   $$\texttt{Breeze}_{1,1} = \texttt{Breeze}_{1,3} = \bot.$$

   He is considering whether to explore the square at $(2, 4)$. He will do so if the probability that it contains a pit is less than 0.4. Should he?

   Hint: The RVs involved are $\texttt{Breeze}_{1,2}, \texttt{Breeze}_{2,3}, \texttt{Breeze}_{1,1}, \texttt{Breeze}_{1,3}$ and $\texttt{Pit}_{i,j}$ for all the (i, j). You need to calculate

   $$\Pr\left(\texttt{Pit}_{2,4} | \text{all the evidence you have so far}\right).$$

2 Maximum likelihood and MAP

1. Several exercises in the problem sheet for *Artificial Intelligence I* last year are relevant to the initial lectures of this course. If you did not attempt them last year, then attempt them now.

2. **Lecture notes slide 49:** Complete the derivation of the MAP learning algorithm for regression

$$\mathbf{w}_{\text{opt}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left[\frac{1}{2} \sum_{i=1}^{m} \left((y_i - h_{\mathbf{w}}(\mathbf{x}_i))^2 \right) + \frac{\lambda}{2} ||\mathbf{w}||^2 \right].$$

3. **Lecture notes slide 56:** Derive the maximum likelihood and MAP algorithms for classification.

3 Linear regression and classification

1. Show that if $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric then

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

   What is the corresponding result when $\mathbf{A}$ is not symmetric?

2. **Lecture notes slide 80:** Show that the optimum weight vector for *ridge regression* is

$$\mathbf{w}_{\text{opt}} = (\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{I})^{-1} \mathbf{\Phi}^T \mathbf{y}.$$

3. Show that if $\mathbf{A} \in \mathbb{R}^{n \times n}$ then

$$\mathbf{A}^T \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{bmatrix} \mathbf{A} = \mathbf{C}$$

   where

$$c_{ij} = \sum_{k=1}^{n} b_k a_{ki} a_{kj}.$$

4. **Lecture notes slide 87:** Show that the Hessian matrix for iterative re-weighted least squares is

$$\mathbf{H}(\mathbf{w}) = \mathbf{\Phi}^T \mathbf{Z} \mathbf{\Phi}.$$

   Hint: you'll need the previous result.

4 Old exam questions

Although this is a new course it has some level of overlap with its predecessor *Artificial Intelligence II*. In particular it might be worth attempting 2010, paper 8, question 2. Also, some old exam questions for *Artificial Intelligence I* are usable warm-ups for the start of this course, so you may like to attempt:

- 2015, paper 4, question 1.
- 2013, paper 4, question 2.
- 2011, paper 4, question 1.
- 2007, paper 4, question 7.