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 $\text{Pit}_{i,j}$ 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 $(i, j)$

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

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

![Wumpus World Grid]

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

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

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 $\text{Breeze}_{1,2}, \text{Breeze}_{2,3}, \text{Breeze}_{1,1}, \text{Breeze}_{1,3}$ and $\text{Pit}_{i,j}$ for all the $(i, j)$. You need to calculate

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

   \[ w_{opt} = \operatorname{argmin}_w \left[ \frac{1}{2} \sum_{i=1}^{m} \left( (y_i - h_w(x_i))^2 \right) + \frac{\lambda}{2} ||w||^2 \right]. \]

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

3 Linear regression and classification

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

   \[ \frac{\partial x^T A x}{\partial x} = 2A x. \]

   What is the corresponding result when \( A \) is not symmetric?

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

   \[ w_{opt} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T y. \]

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

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

   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

   \[ H(w) = \Phi^T Z \Phi. \]

   *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.