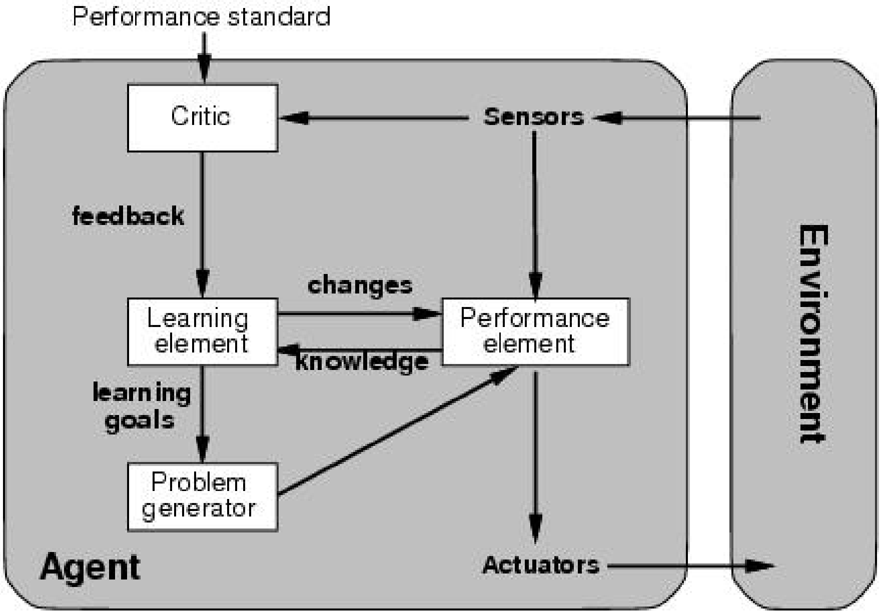
COS30019 - Introduction to Artificial Intelligence
Tutorial Week 11 - Solution

Task 1: Consider the problem faced by an infant learning to speak and understand a language. Explain how this process fits into the general learning model. Describe the percepts and actions of the infant, and the types of learning the infant must do. Describe the subfunctions the infant is trying to learn in terms of inputs and outputs, and available example data.

General learning model:

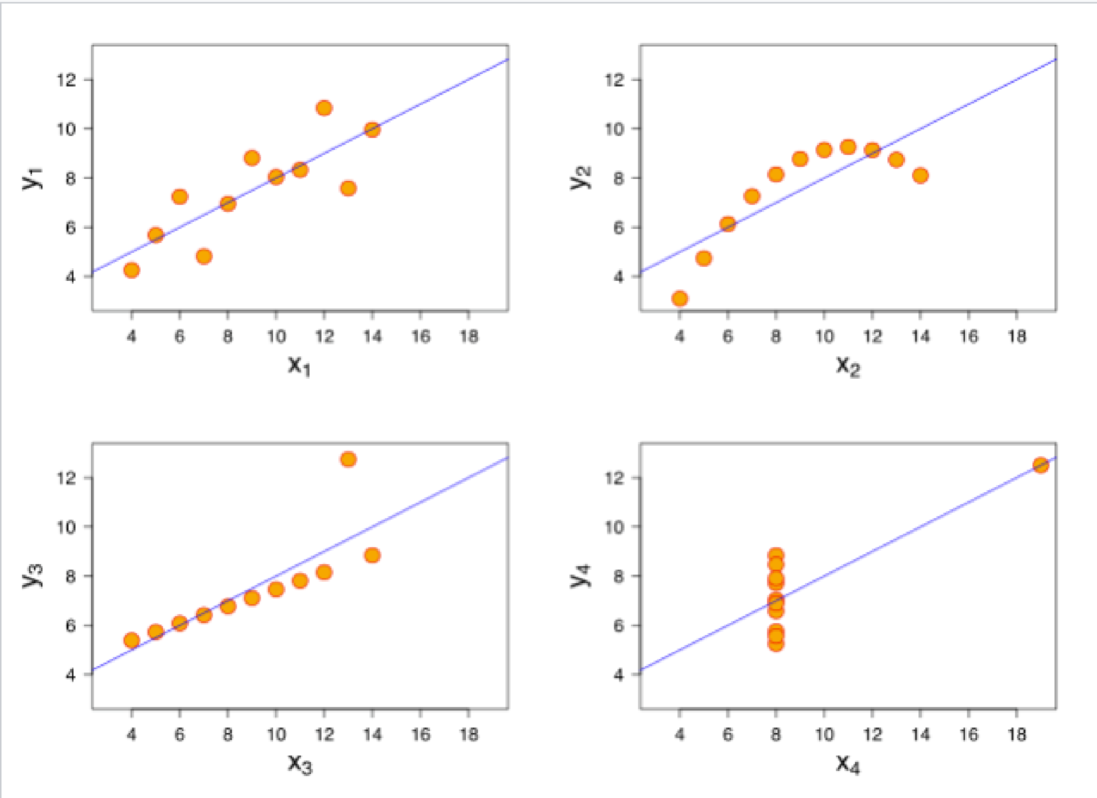


The child is situated in an environment with objects (food, toys, other humans, …) having sensors including ears, eyes, nose, mouth (in fact, the whole body) to allow them to receive percepts of different types (visual, auditory, tactile, etc.) and actuators including mouth, head, body to allow them to perform actions such as making a sound/speech, body gestures (as different forms of communication). The '**P.E.**' takes in the input percepts and is driven by the objectives (food, care, play) and output the actions to request for food, care and play (calling "ma", "pa", "mum mum", "yum yum", pointing to toys, crying, etc.) Subsequently, the learning of the language in a child is part of the learning process the child needs to acquire the knowledge and skills to allow them to maximize their performance measure (of achieving their objectives), for instance: a fruit they are given is a kind of food that can satisfy their hunger and can be referred to as 'apple ' in English. Regarding the task of learning the language, the '**Critic**' component within the child incorporates the percepts from their sensors (pointing to a pear, calling it 'apple') with the external '**Performance Standard**' (e.g., negative reinforcement from a parent by giving corrections: 'No', 'It is a PEAR'). These collective inputs allow the '**Critic**' component to send feedback to the '**L.E**'. The '**L.E.**' then updates the '**P.E.**' to ensure that the child receives the maximum reward (approval/attention from their carer, the food/toys they want, etc.) While it is perhaps less likely that a very young child would use a '**Problem Generator**' to add additional learning goals to the '**P.E.**' to proactively learn additional vocabulary or other language constructs, this component will become more important as the child gets older and seek to enlarge their vocabulary (e.g., to be able to understand a book they are reading or to do a school homework) as well as phonology and grammar.

Task 2: Apart from evaluating the performance (e.g., accuracy) of a linear regression model using a test data set, are there any simple and common techniques that can be used for evaluating a linear regression model (especially for one-dimension input)?

It is generally a good idea to visualize the data when performing linear regression to identify a best-fit line (or, curve). Numerically fitting a line is easy with pure numerical analysis or methods, but determining whether these fitted lines make any sense requires further analysis of which visualizing the data is one of the easiest ways.

The well-known Anscombe's quartet (1973) has typically used to illustrate this point. These are four datasets with almost identical simple statistical properties but appear wildly different in distribution when plotted:



All four sets are identical when examined using simple summary statistics, but vary considerably when graphed

(source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet)

Task 3: Are linear regression models always be visualized as a straight line? Can they model arbitrary curves?

Not all arbitrary curves but any curves that can be expressed by a mathematical function of the input variables can be learned, not just a straight line. For instance, to learn other curves when there is a single input variable x, we can introduce other variables:

$x_1 = x$,
$x_2 = x^2$,
$x_3 = x^3$,
$x_4 = x^{1/2}$,
$x_5 = \log x$,
$x_6 = x^{-0.75}$,
…

And then the linear regression would learn the following function:

$y = h_\beta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \ldots$

Task 4: Discuss the following concepts that have been used extensively in machine learning:

1. Parameters vs. hyperparameters
2. Training data sets vs validation data sets vs test data sets.
3. Training errors vs validation errors vs test errors
4. Overfitting vs. Underfitting

Are there any relations between them? Provide examples in the context of linear regression.

Parameters are internal within a model (and training is the process to determine the values of the parameters to get the 'best-possible' prediction/decision (for instance, in Task 3, β_0, β_1, β_2, β_3, β_4 ,... are the parameters of the linear regression model). Hyperparameters are the explicitly specified parameters that control the training process (such as the learning rates α, number of epochs when training a model, the batch size to be used when training, etc. see this link for number of epochs and batch size: https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/)

If you use ALL data you have to train a model, you run the risk that your model will not be generalizable: It fits very well the data you use for training but does poorly on unseen data (which is what you want your model for) – a phenomenon called *overfitting*. For that reason, it's better to split the data you have into different data sets (*training data sets* – train your model vs *validation data sets* - improve selection of 'good' hyperparameters and adjust your model vs *test data sets* – measure the performance of your model to provide you with confidence when apply/deploy your model).

Task 5: In data science (DS) and machine learning (ML), two important relations are extensively studied: correlation and causation. Are they related? Which relation should typically be modeled by a linear regression model?
Based on this relation, can you propose an approach for agents who operate in uncertain environments that can perform probabilistic reasoning (presented in our **lecture Week 10**)?

Causation (e.g., *Cavity* causes *Toothache*) should imply Correlation but not the other way around (for instance there may be a high correlation between *sales of plants at nurseries* and a *software developer's income*, it is a bad idea to make a career choice based on an increase/decrease in plant sales at Bunnings). With linear regression, we try to model a dependent variable y as a function of the independent variable(s) x; thus, there should be a causation relation between them. Subsequently, a linear regression model can be used to predict the conditional probability between causation-related propositions.