* Chain rule multiple events
* Bayes rule
* Univariate distribution:
  + Central limit theorem – under certain conditions, the result of adding together many random outcomes is approximately gaussian distributed.
  + Easily differentiable
  + Multiply to gaussian
  + Precision is one over variance.
* Multivariate distribution
  + The multivariate generalisation of variance is covariance
* The maximum a posteriori estimation removes the need to calculate the evidence. It is the value of the parameters that maximises the posterior.
  + It is not a distribution but a point value - the mode. Example of bimodal distribution:
* The MLE is the value of theta that maximises the likelihood (i.e. subcase of MAP with uniform priors), where the likelihood is the probability of the observed data given the parameter. It is a means of estimating unknown parameters for the model after observing the data.
* Linear regression:
  + Least squares error is a means of calculating our parameters.
  + Basis functions
  + Data is noisy and even if it wasn’t not going to fit on our basis functions.
  + Regularisation term – essentially a term that penalises overfitting. One of the most common is L2 or ridge that was in the notes.
  + In Bayesian approach – related to which prior chosen.
* Conjugate distributions: when we cannot calculate the evidence we choose a prior that helps us easily calculate the posterior.
  + They give us an exact posterior, easy for online learning. If the posterior and the prior belong to the same distribution family then they are called conjugate distributions.
  + The gaussian family are conjugate to itself – i.e. if the likelihood is gaussian, choosing a gaussian prior over the mean will ensure the posterior distribution is also gaussian.
  + These are chosen for mathematical convenience.
  + Good for online learning.
* Gamma prior conjugate to the Poisson distribution
* Beta prior conjugate to Bernoulli distribution