Chapter 6: Overfitting, Regularization, and Information Criteria

Make a note of the followings:

| Topic | Description | Comment |
| --- | --- | --- |
| Information entropy |  | p stands for a probability distribution of an event. |
| Divergence |  | difference in two entropies.  average difference in log-probabilities of two probability distributions |
| Deviance | an assessment of predictive accuracy  qi is a likelihood (i.e. dnorm, dbinom,..) of each observation. | So the deviance is -2 x sum of log-likelihood of observations  Deviance always declines (i.e. model accuracy erroneously improves) with the number of parameters. This indicates a warning, as unimportant parameters can still improve the accuracy of the model. |
| Flat priors |  | very weak flat priors. The left prior implies the intercept can take any values. This indicates that this prior has no practical effect on inference of a model. |
| Regularizing priors | also known as “skeptical” priors to prevent overfitting (i.e. over-excited by samples)  This means to tell a model to be skeptical of parameter values (in this case slope) above and below 2. In other words, a change in one standard deviation in x units is very unlikely to produce 2 unit changes in the outcome. | **When sample size is small (e.g., N=20), it is important to tune regularizing priors.** Deviance out-of-sample (testing sample) may greatly diverge from deviance in-sample (testing sample).  With sufficient sample size (e.g., N=100), the difference in priors do not make as much a difference as that with small sample size. |
| WAIC (Widely Applicable Information Criterion) | Estimates of deviance (i.e., expected deviance of a model on future data) which can be used to measure which models are better by comparing different WAIC values. | Note that information criteria via DIC, WAIC, or AIC simply estimates accuracy of a model. |
| pWAIC | measures how flexible each model is |  |
| Weight | Akaike weight, each weight can be interpreted as an estimated possibility that each model will perform best on future data. | an estimate of the probability that the model will make the best predictions on new data, conditional on the set of models considered.  For example, **if a model displays 0.97 weight relative to the other models, this means that this model has 97% of the WAIC-estimated model weight.** |