Solving Bayesian Inference

How to get the posterior?

- analytical solution / closed-form solution

$$\text{Beta as prior, Binomial as likelihood -> Beta}$$
$$\text{Gaussian as prior, Gaussian as likelihood -> Gaussian}$$

- numerical solution / approximate solution
- direct sampling from distribution

  For example, Metropolis Hastings algorithm from MCMC family can sample the posterior given a function f whose value is proportional to the posterior. Incidentally, we know that,

$$f(x) = \text{prior}(x)\text{likelihood}(x) \propto \text{posterior}(x)$$

Hypothesis Testing

Hypothesis is an assertion or a statement (about parameters). Hypothesis testing is about telling the significance of the statement.

If you can give a good interval estimation on the parameter, then the significance is easy to derive.

Neyman-Pearson Method

Establish the Hypothesis

null hypothesis H_0, alternative hypothesis H_1. H_0 and H_1 are disjoint.

$$H_0 : \theta = \theta_0$$

1. $H_1 : \theta \neq \theta_0$, double-sided test
2. $H_1 : \theta > \theta_0$, single-sided test
3. $H_1 : \theta < \theta_0$, single-sided test

Establish the Rejection Region

If samples belong to rejection region W, we reject the null hypothesis.

Let $\hat{\theta}$ be a point estimator of θ. Let $c > 0$. When H_1 is double-sided test.

$$W = \{(x_1, \ldots, x_n) || \hat{\theta} - \theta_0| > c\}$$

When $H_1 : \theta > \theta_0$,

$$W = \{(x_1, \ldots, x_n) | \hat{\theta} - \theta_0 > c\}$$

When $H_1 : \theta < \theta_0$,

$$W = \{(x_1, \ldots, x_n) | \hat{\theta} - \theta_0 < c\}$$

Set up the Significance Level

	Judge: Accept H_0	**Judge: Reject H_0**
Fact: H_0 holds	Correct	Type I Mistake
Fact: H_1 holds	Type II Mistake	Correct

We cannot reduce the number of Type I and Type II mistakes at the same time. Usually we choose to reduce the number of Type I mistakes. The chance of Type I mistake is a probability:

$$P((x_1, \ldots, x_n) \in W; \theta = \theta_0)$$

We want this probability to be smaller than a pre-set value α, which is called the significance level. Suppose we are doing double-sided test.

$$W = \{(x_1, \ldots, x_n) || \hat{\theta} - \theta_0| > c\}$$

The significance level is the believed threshold probability of an improbable event.

Supervised Learning

Bayes Classifier

Bayes classifier is a supervised, generative classification model. In generative model, $p(y), p(x|y)$ are learned. $p(y)$ is a categorical distribution whose best estimation is just the frequencies.

Naive Bayes Classifier

In naive Bayesian classifier, we assume a conditional independence for the random vector x given y.

$$p(\mathrm{x}|y) = p(x_1|y)p(x_2|y)\ldots p(x_n|y)$$

Further, we assume that all $p(x_i|y)$ observes the same kind of distribution (in total, $m \times n$ such distributions, where m is the number of different classes). This distribution can be Bernoulli distribution, multinomial distribution, Gaussian distribution, giving Bernoulli naive Bayes, multinomial naive Bayes, Gaussian naive Bayes.

Logistic Regression

Logistic regression is a supervised, discriminative, binary classification model. In discriminative model, $p(y|x)$ is learned.