Notes on Bayesian non-parametrics

Rene Welch

# Basic stuff

## Exponential families

The first concept are exponential families, which are families with a known form and their value is characterized by sufficient statistics:

In the machine learning setting the functions are known as *potentials* and the function is defined so the density function integrates one.

The parameter space is defined as

The *minimal representation* is made in such a way that the potentials are constant. A couple of important results on the *potentials* are:

* $\frac{\partial \Phi(\theta)}{\partial \theta\_a} = \E[ \phi\_a(x)]$
* $\frac{\partial^2 \Phi(\theta)}{\partial \theta\_a \partial \theta\_b} = \E [\phi\_a(x)\phi\_b(x)] - \E[\phi\_a(x)] \E [\phi\_b(x)]$

## Entropy, information and divergence

*Shannon's entropy* is defined as

this function is concave, continuous and maximal for uniform densities. The *Kullback - Leibler* divergence is defined as:

and an important application of this functions is the mutual info. between two random variable and :

In particular, given a target density and an exponential family, the approximating density that minimizes has canonical parameters choosen to match the expected values of that family's sufficient statistics:

$$
\E\_{\hat{\theta}}[\phi\_a(x)] = \int \phi\_a(x)\tilde{p}(x)dx
$$

(the proof of this result is a direct result of optimizing as a function of )

Additionally if , then the MLE of the canonical parameters conincides with the projection defined above:

## Learning with priors

To defined a *full bayesian model* we need the use of a prior distribution, in the case of exponential fmailies the posterior distribution is gonna have the form:

*Proposition* If (an exponential family) with conjugate prior , then the posterior parameters are updated by he rule:

$$
p(\theta | \bf{x},\lambda) = p(\theta | \lambda^\*)
$$

where and

## Examples of conjugate priors:

A couple of typical examples are:

1 - are multinomial and $$ is Dirichlet, a more simple case of this example is when , we have the Beta - Binomial model

2 - is normal and is normal and is inverse Wishart. In the case of being univariate, then is inverse gamma, and the conjugate is *t*

# Graphical models

Hypergraphs provide a mean of describing probability distributions, in which case