ML Course Notes

October 17, 2015

1 Bayes

1.1 ML of Bernoulli

Differentiate the likelihood, set to zero, and solve:

$$\begin{aligned}
L(p) &= p^H(1-p)^T \\
\log L &= H\log p + T\log(1-p) \\
\frac{d\log L}{dp} &= \frac{H}{p} - \frac{T}{(1-p)} \\
0 &= (1-p)H - pT \\
&= H - p(H+T) \\
p &= \frac{H}{H+T}
\end{aligned}$$

1.2 Posterior of Bernoulli-Beta

Re-arrange likelihood and conjugate prior into new posterior beta:

$$\begin{aligned}
p(D|p)p(p) &= \mathrm{Bern}(D|p)\mathrm{Beta}(p;a,b) \\
&= p^H(1-p)^T \cdot p^{a-1}(1-p)^{(b-1)} \\
&= p^{H+a-1}(1-p)^{T+b-1} \\
&= \mathrm{Beta}(H+a, T+b)
\end{aligned}$$

1.2.1 MAP of Bernoulli

Differentiate the posterior beta, set to zero, and solve:

$$\begin{aligned}
p(D|p) &= p^{H+a-1}(1-p)^{T+b-1} \\
\log p(p|D) &= (H+a-1)\log p + (T+b-1)\log(1-p)
\end{aligned}$$

$$
\begin{aligned}
\frac{d\log p(p|D)}{dp} &= (H+a-1)/p - (T+b-1)/(1-p) \\
0 &= (1-p)(H+a-1) - p(T+b-1) \\
&= H' - p(H'+T') \\
p &= H'/(H'+T') \\
&= \frac{H+a-1}{H+a+T+b-2}
\end{aligned}
$$

Letting $H' = H + a$ and $T' = T + b$.

1.3 Posterior Predictive of Bernoulli

Setup the likelihood and the posterior, and integrate out the unknown parameter:

$$
\begin{aligned}
p(x=1|D) &= \int_0^1 p(x=1|p)p(p|D)dp \\
&= \int_0^1 \mathrm{Bern}(x=1|p)\mathrm{Beta}(p|D)dp \\
&= \int_0^1 p \cdot p^{H+a-1}(1-p)^{T+b-1} \\
&= E_{p(p|D)}(p)
\end{aligned}
$$

Expectation of p's posterior. Mean of beta distribution $E[\mathrm{Beta}(p)] = a/a+b$. Plugging in the sufficient statistics from the posterior, we have (Assuming D contained H heads and T tails):

$$
p(x=1|D) = \frac{H+a}{H+a+T+b}
$$

- Could also use a normalization constant strategy.

1.4 Fitting & Bayes Classifiers

1.4.1 MLE of Gaussian

$$
\begin{aligned}
p(X|\mu,\sigma^2) &= \prod_i^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
\log p(x|\mu,\sigma^2) &= -\frac{1}{2}N\log 2\pi\sigma^2 - \sum_i^N \frac{(x_i-\mu)^2}{2\sigma^2} \\
\frac{dL}{d\mu} &= \sum_i^N (x_i-\mu)/\sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\mu &= \frac{1}{N}\sum_i^N x_i \\
\frac{dL}{d\sigma^2} &= -\frac{1}{2}N\frac{1}{\sigma^2} + \sum_i^N (x_i-\mu)^2/2\sigma^4 \\
\sigma^2 N &= \sum_i^N (x_i-\mu)^2 \\
\sigma^2 &= \frac{1}{N}\sum_i^N (x_i-\mu)^2
\end{aligned}
$$

2 Regression

Task is

$$
\begin{aligned}
\min \quad & \sum_i^N (y_i - f(x_i, w))^2 \\
&= (\mathbf{y} - \mathbf{W}\mathbf{x})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
\frac{dE}{d\mathbf{w}} &= -\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
\mathbf{X}^T\mathbf{y} &= \mathbf{X}^T\mathbf{X}\mathbf{w} \\
\mathbf{w}_{ols} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}
\end{aligned}
$$

OR

$$
\begin{aligned}
\mathbf{w} \quad := \quad & \mathbf{w} - \alpha\frac{dE}{d\mathbf{w}} \\
= \quad & \mathbf{w} + \alpha\left(\sum_i \mathbf{x}_i(\mathbf{y}_i - \mathbf{w}^T\mathbf{x}_i)\right) \\
OR \quad & \mathbf{w} + \alpha\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w})
\end{aligned}
$$