BOOSTING, LOG ODDS, AND BINARY BAYES FILTERS

ALEX TEICHMAN

1. Binary Bayes Filters

In the binary Bayes filter, we wish to estimate the log odds l_T of a binary variable $y \in \{-1, +1\}$ given a series of measurements $z_{1:T}$. The variable y might indicate grid cell occupancy or whether a tracked cluster in Velodyne laser data is a pedestrian, for example. The update rule (see page 94) is

$$l_t = l_{t-1} + \log \tfrac{P(y|z_t)}{P(\neg y|z_t)} - \log \tfrac{p(y)}{p(\neg y)}. \tag{1}$$

2. Boosting as a Log Odds Estimator

The boosting optimization problem is $\min_{H(z)} \mathbb{E}_{P(x,y)} [\exp(-yH(z))]$; that is, we are looking for a function that minimizes exponential loss. Writing out the expectation explicitly and assuming z is discrete (though this also works for real-valued z), we have

$$\begin{aligned}
\mathbb{E}_{P(z,y)} [\exp(-yH(z))] &= \sum_z \sum_y P(z,y) \exp(-yH(z)) && (2)\\
&= \sum_z p(z) \sum_y P(y|z) \exp(-yH(z)) \\
&= \sum_z p(z) \mathbb{E}_{P(y|z)} [\exp(-yH(z))] . && (3)
\end{aligned}$$

This may not seem useful until we examine the conditional expectation. It is convex in $H(z)$, so we know when the derivative with respect to $H(z)$ is zero we are at the minimum.

$$\begin{aligned}
\mathbb{E}_{P(y|z)} [\exp(-yH(z))] &= P(y=1|z) \exp(-H(z)) + P(y=-1|z) \exp(H(z)) \\
\frac{\partial}{\partial H(z)} \mathbb{E}_{P(y|z)} [\exp(-yH(z))] &= -P(y=1|z) \exp(-H(z)) + P(y=-1|z) \exp(H(z))
\end{aligned}$$

Setting the derivative to zero, we have

$$\begin{aligned}
P(y=-1|z) \exp(H(z)) &= P(y=1|z) \exp(-H(z)) \\
\exp(2H(z)) &= \frac{P(y=1|z)}{P(y=-1|z)} \\
H(z) &= \frac{1}{2} \log \frac{P(y=1|z)}{P(y=-1|z)}.
\end{aligned}$$

This shows that the conditional expectation is minimized when the boosting strong classifier $H(z)$ returns the log odds (up to a constant factor of a half). Looking back at (3), the joint expectation (2) is minimized if the conditional expectation for every possible z is minimized. As we just showed this happens when $H(z)$ returns the log odds, it follows that the joint expectation is minimized when $H(z)$ returns the log odds.

Now, we can make the connection to binary Bayes filters: $2H(z)$ can be used for the second term of (1), and we can use the binary Bayes math for integrating our boosting predictions over time.

Date: February 24, 2010.