后验概率估计

- 回顾下最大似然估计
  - 找到最合适的参数使其满足咱们的数据 $\underset{\mu}{\operatorname{argmax}}\, p(\mathbf{X};\mu)$
  - 还记得咱们的最大似然估计吧 $\underset{\mu}{\operatorname{argmax}}\, p(\mathbf{X};\mu) = \underset{\mu}{\operatorname{argmax}}\, \log p(\mathbf{X};\mu)$
  - 如果一件事，正例的出现了8次，负例出现了2次，如何来求解呢?

后验概率估计

- 最大似然估计

求解的目标 $p(\mathbf{X};\mu)=\prod_{i=1}^{n}p(x_i;\mu)=\prod_{i=1}^{n}\mu^{x_i}(1-\mu)^{1-x_i}$

$$
\begin{aligned}
\log p(\mathbf{X};\mu) &= \log\prod_{i=1}^{n}\mu^{x_i}(1-\mu)^{1-x_i} \\
&= \sum_{i=1}^{n}\log\left\{\mu^{x_i}(1-\mu)^{1-x_i}\right\} \\
&= \sum_{i=1}^{n}\left[\log\mu^{x_i}+\log(1-\mu)^{1-x_i}\right] \\
&= \sum_{i=1}^{n}\left[x_i\log\mu+(1-x_i)\log(1-\mu)\right]
\end{aligned}
$$

后验概率估计

最大似然估计

该求导了

$$
\begin{aligned}
\frac{\partial}{\partial \mu}\log p(\mathbf{X};\mu) &= \sum_{i=1}^{n}\frac{\partial}{\partial \mu}[x_i\log\mu+(1-x_i)\log(1-\mu)] \\
&= \sum_{i=1}^{n}x_i\frac{\partial}{\partial \mu}\log\mu+\sum_{i=1}^{n}(1-x_i)\frac{\partial}{\partial \mu}\log(1-\mu) \\
&= \frac{1}{\mu}\sum_{i=1}^{n}x_i-\frac{1}{1-\mu}\sum_{i=1}^{n}(1-x_i)
\end{aligned}
$$

后验概率估计

- 最大后验概率有啥区别吗

要求的东西变了吗？好像木有，都是做参数估计。

问题变得复杂一点了，现在多了一个先验知识。

优化的目标：$\hat{\mu}_{MAP} = \underset{\mu}{\operatorname{argmax}}\, p(\mu|\mathbf{X})$

根据贝叶斯公式：
$$\begin{aligned}\hat{\mu}_{MAP} &= \underset{\mu}{\operatorname{argmax}}\, p(\mu|\mathbf{X}) \\ &= \underset{\mu}{\operatorname{argmax}}\, \frac{p(\mathbf{X}|\mu)p(\mu)}{p(\mathbf{X})} \\ &= \underset{\mu}{\operatorname{argmax}}\, p(\mathbf{X}|\mu)p(\mu)\end{aligned}$$

这两项分别是什么啊？有点眼熟。

后验概率估计

最大后验估计

- $p(\mathbf{X}|\mu)$这不就是我们的似然嘛，$p(\mu)$这就是先验知识了。
- 单如求解呢？照样，但是好像多了个先验，这个你得告诉我吧。

$$
\begin{aligned}
\operatorname*{argmax}_{\mu} \Pr(\mu|\mathbf{X}) &= \operatorname*{argmax}_{\mu} \log \Pr(\mu|\mathbf{X}) \\
&= \operatorname*{argmax}_{\mu} \log \prod_{x_i \in \mathbf{X}} \Pr(x_i|\mu) \cdot \Pr(\mu) \\
&= \operatorname*{argmax}_{\mu} \sum_{x_i \in \mathbf{X}} \{\log \Pr(x_i|\mu)\} + \log \Pr(\mu)
\end{aligned}
$$

后验概率估计

最大后验估计

- 假设参数有一个先验服从Beta分布 $\Pr(\mu) = \mathrm{Beta}(\mu|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \cdot \mu^{\alpha-1}(1-\mu)^{\beta-1}$

- 我们要试验的任务当做一个伯努利实验吧，每次都是一个二项分布

$$\Pr(x_i|\mu) = \mathrm{Bernoulli}(x_i|\mu) = \mu^{x_i}(1-\mu)^{1-x_i}$$

- 目标这就来了：$\frac{\partial}{\partial\mu}\mathcal{L} = \sum_i \frac{\partial}{\partial\mu} \log \mathrm{Bernoulli}(x_i|\mu) + \frac{\partial}{\partial\mu} \log \mathrm{Beta}(\mu|\alpha,\beta)$