3.15 PAC-Bayesian Bounds

Bayesian estimation procedure

(i) prior distribution $P(h)$ over hypotheses.

(ii) training data $z_1, \cdots, z_n$

(iii) F : likelihood function

Produce a posterior function $Q(h) \propto P(h) \prod_{i=1}^{n} F(z_i | h)$

Note: Bayesian procedure assume P and F are correct.

What if P or F is wrong ?

1. Bounds that depends on the prior

Recall that for finite hypothesis H with loss bounded in $[0,1]$, by Hoeffding's inequality : with prob. $\geq 1-\delta$

$$\forall h \in H : \ L(h) \leq \hat{L}(h) + \sqrt{\frac{\log|H| + \log(1/\delta)}{2n}}$$

Note: In this bound, each h is treated the same.

[Theorem 17] Occam bound

Let H be a countable hypothesis class.

Let the loss function be bounded : $\ell(z, h) \in [0,1]$

Let P be any prior distribution over H.

Then with prob. $\geq 1-\delta$,

$$\forall h \in H : \ L(h) \leq \hat{L}(h) + \sqrt{\frac{\log(1/P(h)) + \log(1/\delta)}{2n}}$$

Pf of Theorem 17:

By Hoeffding's ineq., for $\forall h \in H$

$$P\{L(h) \geq \hat{L}(h) + \varepsilon\} \leq \exp(-2n\varepsilon^2)$$

If we set RHS to $\delta P(h)$:

$$L(h) \geq \hat{L}(h) + \sqrt{\log(1/\delta P(h))}$$

$L(h) \geq \hat{L}(h) + \sqrt{\frac{\cdots}{2n}}$

Apply union bound arcoss $\mathcal{H}$: with prob. $\geq 1-\delta$

$$\forall h \in \mathcal{H}: \; L(h) \geq \hat{L}(h) + \sqrt{\frac{\log(1/\delta P(h))}{2n}} \qquad \square$$

Note: (i) This suggest an algorithm: RHS to be the objection and minimise it.

(ii) $A(S) := \underset{h \in R}{\arg\min} \, \hat{L}(h) + R(h)$, $R(h) := \sqrt{\frac{\log(1/P(h)) + \log(1/\delta)}{2n}}$

(iii) $n \to \infty$, $A(S) \to 0$.

(iv) This is not a Bayesian procedure. If we the the loss be $\ell(z,h) = -\log F(z|h)$, then MAP would be $\underset{h \in \mathcal{H}}{\arg\min} \, \hat{L}(h) + \frac{\log(1/P(h))}{n}$ (the only difference is the root)

Lack: (i) $\mathcal{H}$ is countable

(ii) Only embrace half of the Bayesian story: when we have a prior $P(h)$, only a single $h \in \mathcal{H}$ is returned.

[Theorem 18] PAC-Bayesian theorem

Let loss $\ell(z,h) \in [0,1]$

Let P be any prior over $\mathcal{H}$

Let Q_S be any posterior over $\mathcal{H}$ (a function of S)

Then with prob. $\geq 1-\delta$

$$E_{h \sim Q(S)}[L(h)] \leq E_{h \sim Q(S)}[\hat{L}(h)] + \sqrt{\frac{KL(Q_S \| P) + \log(4n/\delta)}{2n-1}}$$

Note: set Q_S be a point mass at some h, we get Th. 16.
(up to constants)