Compare Classifiers (Pt 2)

10/21

Given 2 classifiers M^A & M^B
which is better on D

1. use K-fold cross validation on M^A & M^B
→ $\theta_1^A, \theta_2^A, \dots, \theta_K^A$ ← performance of M^A
$\theta_1^B, \theta_2^B, \dots, \theta_K^B$ ← performance of M^B

2. define $\delta_i = \theta_i^A - \theta_i^B$ (diff of performance on fold i)

3. consider expected diff & var of diff

$$\hat{\mu}_\delta = \frac{1}{K}\sum_{i=1}^{K} \delta_i$$

$$\hat{\sigma}_\delta^2 = \frac{1}{K}\sum_{i=1}^{K} (\delta_i - \hat{\mu}_i)^2$$

4. Use hypothesis testing

a. null hyp H_0: $\mu_\delta = 0$

alternative hyp H_a: $\mu_\delta \neq 0$

6. define z-score $Z_\delta^* = \sqrt{K}\left(\frac{\hat{\mu}_\delta - \mu}{\hat{\sigma}_\delta}\right)$ ($\mu = 0$ under null hyp) $\rightarrow Z_\delta^* = \frac{\sqrt{K}\,\hat{\mu}_\delta}{\hat{\sigma}_\delta}$

C. given significale level α ← we want this value before comparison

for single test typically $\alpha = .05$

we have $P(Z_j^* \in [-t_{\alpha/2}, t_{\alpha/2}]) = 1-\alpha$

↖ quantile function of student's t-dist w/ prob $1-\alpha/2$

$\Rightarrow$ if $Z_j^* \notin (-t_{\alpha/2}, t_{\alpha/2})$ $\Rightarrow$ we can reject the H_0

Otherwise $Z_j^* \in [-t_{\alpha/2}, t_{\alpha/2}]$

fail to reject the null hyp

Ensamble classifiers

"if 1 is good are 50 better?"

Classifiers is **unstable** if a small perturbation of the training set produces a large change in prediction

reduce instability:
use **ensamble** to create a combine classifier from a set of base-classifers,
↳ collection of classifers
each is trained on a diff subset of data

Bagging (Bootstrap AGGrig~~ating~~)

Given $D = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ w/ $\vec{x}_i \in \mathbb{R}^d$

$K \in \mathbb{Z}^+$

Create D_t $t = 1, \dots, K$
each D_t is a different training set
created w/ bootstrap sampling: sample from D w/ replacement

Learn M_t on D_t

Classify $\vec{x}$ w/ majority vote

$$v_j(\vec{x}) = \sum_{i=1}^{K} I(M_i(\vec{x}) = c_j)$$

← # of times classifier predicts c_j

pick maximal w/ combined classifier

$$M^K(\vec{x}) = \underset{c_j}{\operatorname{argmax}} \{ v_j(\vec{x}) \mid j = 1, \dots, K \}$$

← majority label