Expected in-sample error and optimism

Alessandro Chiuso

November 16, 2016
In-sample error and training errors

Definitions

Let us consider the new (not used for training) data

\[ y_i^0 := f(x_i) + e_i^0 \quad i = 1, \ldots, N \]

where \( x_i \) are the input locations used in the training set while \( e_i^0 \) is a “new” error independent of training data \( y_i \)

Consider the in-sample error

\[ Err_{in} := \frac{1}{N} \sum_{i=1}^{N} E_{Y_0}[L(y_i^0, \hat{f}(x_i)|\tau] \]

where \( \tau := \{(x_i, y_i)\}_{i=1,\ldots,N} \) is the training data set. Let us also define the training error

\[ \overline{err} = E_{\text{train}} := \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{f}(x_i)) \]
Optimism and expected

We have defined the optimism

\[ OPT := Err_{in} - E_{train} \]

The following theorem holds:

**Theorem**

\[ \omega := EOPT = \frac{2}{N} \sum_{i=1}^{N} cov \{ \hat{y}_i, y_i \} \]

where \( \hat{y}_i := \hat{f}(x_i) \) and \( \hat{f} \) has been estimated using (and thus is a function of) the training data \( \tau \).
Expected optimism - proof (I)

We shall provide the proof for the case of squared loss 
\[ L(y, f(x)) := (y - f(x))^2. \]

**Proof:**
Let us define \( Y := [y_1, ..., y_N]^\top \), \( \hat{F}(Y) := [\hat{f}(x_1), ..., \hat{f}(x_N)]^\top \), 
\( Y^0 := [y_1^0, ..., y_N^0]^\top \) and similarly \( E \) and \( E^0 \). Then

\[ E_{\text{train}} = \frac{1}{N} \| Y - \hat{F}(Y) \|^2 \]

and

\[ \text{Err}_{\text{in}} = \frac{1}{N} \mathbb{E}_{Y^0} [\| Y^0 - \hat{F}(Y) \|^2 | \tau ] \]

Then, defining \( F := [f(x_1), .., f(x_N)]^\top \) (the “true” regression function) we have

\[ Y^0 = F + E^0 \quad Y^0 = F + E^0 \]

so that

\[ Y^0 - \hat{F}(Y) = Y - E^0 + E - \hat{F}(Y) \]
Expected optimism - proof (II)

It thus follows that

\[ \| Y^0 - \hat{F}(Y) \|^2 = \| Y - \hat{F}(Y) \|^2 + \| E - E^0 \|^2 + \]

\[ = \underbrace{\| Y - \hat{F}(Y) \|^2}_{(A)} + \underbrace{\| E - E^0 \|^2}_{(B)} + \]

\[ + 2 \underbrace{(Y - \hat{F}(Y))^\top (E^0 - E)}_{(C)} \] (1)

Now, the following equalities hold:

\[ \mathbb{E}[(A)] = N \mathbb{E} \mathbb{E}_{train} \]

Using the fact that \( E \) and \( E^0 \) are independent, zero mean and with variances \( \text{Var}\{E\} = \text{Var}\{E^0\} = \sigma^2 I \),

\[ \mathbb{E}[(B)] = 2N\sigma^2 \]
Expected optimism - proof (III)

\[ E[(C)] = E[(Y - \hat{F}(Y))^\top (E^0 - E)] = \]
\[ = \mathbf{E}[(Y - \hat{F}(Y))E^0 - \mathbf{E}[(Y - \hat{F}(Y))^\top E]] \]
\[ = -\mathbf{E}[(F - \hat{F}(Y))^\top E] - \mathbf{E}[E^\top E] \]
\[ = \sum_{i=1}^{N} \mathbf{E}[(f(x_i) - \hat{f}(x_i))e_i] - N\sigma^2 \]
\[ = \sum_{i=1}^{N} \text{cov}(\hat{f}(x_i), y_i) - N\sigma^2 \]

Inserting these expressions for (expected values of) the terms (A), (B) and (C) in (1) we have:

\[ E\|Y^0 - \hat{F}(Y)\|^2 = NEE_{\text{train}} + 2N\sigma^2 - 2N\sigma^2 + 2 \sum_{i=1}^{N} \text{cov}(\hat{f}(x_i), y_i) \]

Thus leading to

\[ E[Err_{in}] - E[E_{\text{train}}] = \frac{1}{N} E\|Y^0 - \hat{F}(Y)\|^2 - E E_{\text{train}} \]
\[ = \frac{2}{N} \sum_{i=1}^{N} \text{cov}(\hat{f}(x_i), y_i) \]

which concludes the proof