Caracterización del error de un clasificador

Percepción - ETSInf

Se parte de definir el clasificador G, entrenado a partir de muestras de entrenamiento, como un regresor, de forma que:

$$G(x) : E \to \mathbb{R}$$

Se define también el valor verdadero para $x \in E$ como un valor $y \in \mathbb{R}$ tal que:

$$y = F(x) + \epsilon$$

Donde F sería la función verdadera que se trata de estimar con G y ϵ es el ruido inherente a los datos.

Dado esto, se define el criterio de error como el valor esperado (esperanza matemética $\mathbb{E}$) del error cuadrático:

$$\mathbb{E}[(y - G(x))^2]$$

Desarrollando esto, tenemos:

$$\mathbb{E}\left[(y - G(x))^2\right] = \mathbb{E}\left[y^2 - 2\,y\,G(x) + G(x)^2\right] =$$

$$\mathbb{E}\left[G(x)^2\right] - 2\ \mathbb{E}[G(x)]\ \mathbb{E}[y] + \mathbb{E}\left[y^2\right]$$

Una propiedad de la esperanza matemática $\mathbb{E}$ es que, sobre cualquier variable aleatoria Z, se tiene:

$$\mathbb{E}[Z^2] = \mathbb{E}\left[(Z - \mathbb{E}\,[Z])^2\right] + \mathbb{E}[Z]^2$$

Llamando $\overline{Z} = \mathbb{E}[Z]$, esto quedaría como:

$$\mathbb{E}[Z^2] = \mathbb{E}\left[\left(Z - \overline{Z}\right)^2\right] + \overline{Z}^2$$

Por tanto, llamando también $\overline{G(x)} = \mathbb{E}\,[G(x)]$ y $\overline{y} = \mathbb{E}[y]$, tendremos que:

$$\mathbb{E}[(y - G(x))^2] = \mathbb{E}\left[\left(G(x) - \overline{G(x)}\right)^2\right] + \overline{G(x)}^2 - 2\,\overline{G(x)}\,\overline{y} + \mathbb{E}\left[(y - \overline{y})^2\right] + \overline{y}^2$$

Aquí se puede aplicar la propiedad de cancelación del error, que nos dice que $\overline{y} = \mathbb{E}\left[F(x) + \epsilon\right] = F(x)$, con lo que:

$$\mathbb{E}[(y - G(x))^2] = \mathbb{E}\left[\left(G(x) - \overline{G(x)}\right)^2\right] + \overline{G(x)}^2 - 2\,\overline{G(x)}\,F(x) + \mathbb{E}\left[(y - F(x))^2\right] + F(x)^2 =$$

$$\mathbb{E}\left[\left(G(x) - \overline{G(x)}\right)^2\right] + \mathbb{E}\left[(y - F(x))^2\right] + \overline{G(x)}^2 - 2\,\overline{G(x)}\,F(x) + F(x)^2 =$$

$$\mathbb{E}\left[\left(G(x) - \overline{G(x)}\right)^2\right] + \mathbb{E}\left[(y - F(x))^2\right] + \left(\overline{G(x)} - F(x)\right)^2$$

Cada uno de estos términos representa lo siguiente:

- $\mathbb{E}\left[\left(G(x) - \overline{G(x)}\right)^2\right]$ es la *variance*: sensibilidad del clasificador al entrenamiento disponible, es decir, capacidad de generalización del clasificador
- $\left(\overline{G(x)} - F(x)\right)^2$ es el *bias*: capacidad del clasificador de ajustarse a la auténtica distribución de los datos
- $\mathbb{E}\left[(y - F(x))^2\right]$ es el *noise*: ruido presente en los datos