Học máy và Thị giác máy tính

Đinh Viết Sang

December 20, 2022

Mục lục

Chương 1

Bias và Variance

1.1 Sai số kiểm tra kỳ vọng (expected test error)

Xét bài toán học có giám sát với tập học $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ được lấy mẫu độc lập và tuân theo phân phối giống nhau (*i.i.d.*) từ phân phối dữ liệu $\Omega(\mathcal{X}, \mathcal{Y})$, trong đó $\mathcal{X}$ là không gian dữ liệu và $\mathcal{Y}$ là không gian nhãn.

Giả sử $f : \mathcal{X} \to \mathcal{Y}$ là ánh xạ chính xác biểu diễn quan hệ từ một dữ liệu đầu vào $x \in \mathcal{X}$ sang một nhãn đầu ra $y \in \mathcal{Y}$:

$$y = f(\mathbf{x}) + \epsilon \tag{1.1}$$

trong đó $\epsilon = \mathcal{N}(0, \beta^2)$ là nhiễu Gauss với kỳ vọng bằng 0.

Thông thường nhãn y không thể thu thập chính xác được do sai số khách quan trong quá trình thu thập dữ liệu liệu. Chẳng hạn, khi xét bài toán hồi quy cân nặng, chiều cao, các chỉ số sinh hoá, giá trị nhãn y có thể bị sai lệch do thiết bị đo đạc và thiết bị xét nghiệm... Trong bài toán hồi quy giá nhà, cùng một $\mathbf{x}$ (cùng diện tích, cùng tầng, cùng số phòng...) nhưng có thể có nhiều mức giá y khác nhau tuỳ thời điểm và các yếu tố khách quan khác. Kỳ vọng của nhãn y thu được ứng với dữ liệu đầu vào x chính là $f(\mathbf{x})$:

$$f(\mathbf{x}) = \mathbb{E}\big[y|\mathbf{x}\big] = \int_y y Pr(y|\mathbf{x}) \partial y \tag{1.2}$$

Giả sử ta xấp xỉ $f(\mathbf{x})$ bằng giả thuyết $h(\mathbf{x}, \mathbf{D})$ là một mô hình có tham số được huấn luyện trên tập học D. Quá trình huấn luyện h thường được quy về bài toán cực tiểu hóa hàm rủi ro thực nghiệm (emprirical risk hoặc training error):

$$\hat{R}_D(h) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(h(\mathbf{x}_i, D), y_i) \tag{1.3}$$

trong đó $\mathcal{L}(h(\mathbf{x}_i, D), y_i)$ là hàm chi phí được định nghĩa trước.

Tuy nhiên, hiệu năng của mô hình phải được đánh giá trên toàn bộ phân phối dữ liệu $\mathcal{D}$, bao gồm cả những dữ liệu chưa nhìn thấy (dữ liệu test), sử dụng độ đo gọi là rủi ro kỳ vọng (expected risk hoặc test error):

$$R_D(h) \triangleq \mathbb{E}_{(\mathbf{x},y)\sim\Omega}\big[\mathcal{L}(h(\mathbf{x}, D), y)\big] = \int_{\mathbf{x}} \int_y \mathcal{L}(h(\mathbf{x}, D), y) Pr(\mathbf{x}, y) \partial y \partial \mathbf{x} \tag{1.4}$$

Giả thuyết h khi huấn luyện trên các tập học D khác nhau sẽ thu được các mô hình $h(\mathbf{x}, D)$ khác nhau ứng với các bộ tham số khác nhau. Do đó khi đánh giá giả thuyết h, ta

quan tâm tới sai số kiểm tra kỳ vọng (expected test error) ứng với mọi trường hợp khác nhau của tập học D:

$$\mathbb{E}_D\big[R_D(h)\big] = \int_D R_D(h)Pr(D)\partial D = \int_D \int_{\mathbf{x}} \int_y \mathcal{L}(h(\mathbf{x}, D), y)Pr(\mathbf{x}, y)Pr(D)\partial y \partial \mathbf{x} \partial D$$
$$= \mathbb{E}_{\mathbf{x},y,D}\big[\mathcal{L}(h(\mathbf{x}, D), y)\big] \tag{1.5}$$

Mô hình kỳ vọng $\bar{h}(\mathbf{x})$ là trung bình tất cả các mô hình $h(\mathbf{x}, D)$ được huấn luyện trên tất cả các tập học D khác nhau:

$$\bar{h}(\mathbf{x}) = \mathbb{E}_D\big[h(\mathbf{x}, D)\big] = \int_D h(\mathbf{x}, D)Pr(D)\partial D \tag{1.6}$$

1.2 Phân rã bias và variance

Xét trường hợp hồi quy tuyến tính với hàm mất mát bình phương $\mathcal{L}(h(\mathbf{x}, D), y) = (h(\mathbf{x}, D) - y)^2$. Khi đó sai số kiểm tra kỳ vọng có thể khai triển như sau:

$$\begin{aligned}
\mathbb{E}_{\mathbf{x},y,D}\big[\mathcal{L}(h(\mathbf{x}, D), y)\big] &= \mathbb{E}_{\mathbf{x},y,D}\big[\big(h(\mathbf{x}, D) - y\big)^2\big] = \mathbb{E}_{\mathbf{x},y,D}\big[\big((h(\mathbf{x}, D) - \bar{h}(\mathbf{x})) + (\bar{h}(\mathbf{x}) - y)\big)^2\big] \\
&= \mathbb{E}_{\mathbf{x},y,D}\big[\big(h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big)^2\big] + \mathbb{E}_{\mathbf{x},y,D}\big[\big(\bar{h}(\mathbf{x}) - y\big)^2\big] \\
&+ 2\mathbb{E}_{\mathbf{x},y,D}\big[\big(h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big)\big(\bar{h}(\mathbf{x}) - y\big)\big] \\
&= \mathbb{E}_{\mathbf{x},D}\big[\big(h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big)^2\big] + \mathbb{E}_{\mathbf{x},y}\big[\big(\bar{h}(\mathbf{x}) - y\big)^2\big] \\
&+ 2\mathbb{E}_{\mathbf{x},y,D}\big[\big(h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big)\big(\bar{h}(\mathbf{x}) - y\big)\big]
\end{aligned} \tag{1.7}$$

Xét số hạng thứ 3 trong Công thức (1.7), do $\bar{h}(\mathbf{x}) - y$ không phụ thuộc vào D nên có thể đặt làm thừa số chung:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x},y,D}\Big[\big(h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big)\big(\bar{h}(\mathbf{x}) - y\big)\Big] = \mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - y\big)\mathbb{E}_D\big[h(\mathbf{x}, D) - \bar{h}(\mathbf{x})\big]\Big] \\
&= \mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - y\big)\Big(\mathbb{E}_D\big[h(\mathbf{x}, D)\big] - \mathbb{E}_D\big[\bar{h}(\mathbf{x})\big]\Big)\Big] \\
&= \mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - y\big)\Big(\bar{h}(\mathbf{x}) - \bar{h}(\mathbf{x})\Big)\Big] \\
&= \mathbb{E}_{\mathbf{x},y}\Big[0\Big] \\
&= 0
\end{aligned} \tag{1.8}$$

Xét số hạng thứ 2 của Công thức (1.7), ta có:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - y\big)^2\Big] = \mathbb{E}_{\mathbf{x},y}\Big[\big((\bar{h}(\mathbf{x}) - f(\mathbf{x})) + (f(\mathbf{x}) - y)\big)^2\Big] \\
&= \mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)^2\Big] + \mathbb{E}_{\mathbf{x},y}\Big[(f(\mathbf{x}) - y)^2\Big] + 2\mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\big(f(\mathbf{x}) - y\big)\Big] \\
&= \mathbb{E}_{\mathbf{x}}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)^2\Big] + \mathbb{E}_{\mathbf{x},y}\Big[(f(\mathbf{x}) - y)^2\Big] + 2\mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\big(f(\mathbf{x}) - y\big)\Big]
\end{aligned} \tag{1.9}$$

Xét số hạng thứ 3 trong Công thức (1.9), do $\bar{h}(\mathbf{x}) - f(\mathbf{x})$ không phụ thuộc vào y nên có thể đặt làm thừa số chung như sau:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x},y}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\big(f(\mathbf{x}) - y\big)\Big] = \mathbb{E}_x\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\mathbb{E}_{y|x}\big[f(\mathbf{x}) - y\big]\Big] \\
&= \mathbb{E}_x\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\big(\mathbb{E}_{y|x}\big[f(\mathbf{x})\big] - \mathbb{E}_{y|x}\big[y\big]\big)\Big] \\
&= \mathbb{E}_x\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)\big(f(\mathbf{x}) - f(\mathbf{x})\big)\Big] \\
&= 0
\end{aligned} \tag{1.10}$$

Kết hợp các Công thức (1.7), (1.8), (1.9), (1.10), ta có phân rã cuối cùng:

$$\underbrace{\mathbb{E}_{\mathbf{x},y,D}\big[\big(h(\mathbf{x},D) - y\big)^2\big]}_{expected\ test\ error} = \underbrace{\mathbb{E}_{\mathbf{x},D}\big[\big(h(\mathbf{x},D) - \bar{h}(\mathbf{x})\big)^2\big]}_{variance} + \underbrace{\mathbb{E}_{\mathbf{x}}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)^2\Big]}_{bias^2} + \underbrace{\mathbb{E}_{\mathbf{x},y}\Big[\big(f(\mathbf{x}) - y\big)^2\Big]}_{noise} \tag{1.11}$$

Như vậy ta có các khái niệm sau:

Bias của giả thuyết h tại một điểm dữ liệu vào $\mathbf{x}$ thể hiện sự chênh lệch giữa giá trị nhãn lý tưởng $f(\mathbf{x})$ so với kỳ vọng $\bar{h}(\mathbf{x})$ của các phán đoán $h(\mathbf{x}, D)$ do các mô hình được huấn luyện trên các tập học D khác nhau đưa ra.

Nếu xét trên toàn bộ không gian dữ liệu $\mathcal{X}$, ta có khái niệm bias bình phương (*squared bias*):

$$bias^2 = \mathbb{E}_{\mathbf{x}}\Big[\big(\bar{h}(\mathbf{x}) - f(\mathbf{x})\big)^2\Big] = \int_{\mathbf{x}}\left(\int_D h(\mathbf{x},D)Pr(D)\partial D - f(\mathbf{x})\right)^2 Pr(\mathbf{x})\partial\mathbf{x} \tag{1.12}$$

Variance của giả thuyết h tại một điểm dữ liệu vào $\mathbf{x}$ thể hiện sự phân tán của các phán đoán khác nhau $h(\mathbf{x}, D)$ do các mô hình được huấn luyện trên các tập học D khác nhau đưa ra.

Nếu xét trên toàn bộ không gian dữ liệu $\mathcal{X}$, ta có khái niệm variance như sau:

$$\begin{aligned}
variance &= \mathbb{E}_{\mathbf{x},D}\big[\big(h(\mathbf{x},D) - \bar{h}(\mathbf{x})\big)^2\big] \\
&= \int_{\mathbf{x}}\int_D\left(h(\mathbf{x},D) - \int_D h(\mathbf{x},D)Pr(D)\partial D\right)^2 Pr(D)Pr(\mathbf{x})\partial D\partial\mathbf{x}
\end{aligned} \tag{1.13}$$