2.2 Giải thuật giảm gradient ngẫu nhiên

Một cài đặt cho giải thuật SVM của (Bottou & Boussquet, 2008), (Shalev-Shwartz *et al*., 2007) dựa trên phương pháp giảm gradient ngẫu nhiên (SGD), có độ phức tạp tuyến tính với số phần tử dữ liệu. Từ các ràng buộc (không xét độ lệch *b*) trong (1), có thể viết lại như sau:

$$z_i \geq 1 - y_i(w.x_i) \qquad (2)$$

$$z_i \geq 0 \ (i=1,m) \qquad (3)$$

Các ràng buộc (2), (3) có thể được viết ngắn gọn như (4):

$$z_i = max\{0, 1 - y_i(w.x_i)\} \qquad (4)$$

Bằng cách thay thế z_i vào hàm mục tiêu của (1), việc tìm siêu phẳng tối ưu của SVM có thể được thực hiện bởi (5):

$$min\ \Psi(w, x, y) = (\lambda/2)\ ||w||^2 + (1/m) \sum_{i=1}^{m} \max\{0, 1 - y_i(w.x_i)\} \qquad (5)$$

Phương pháp giảm gradient (GD) thực hiện tối ưu vấn đề (5) bằng cách cập nhật *w* tại lần lặp thứ *(t+1)* với tốc độ học η_t, như trong (6):

$$w_{t+1} = w_t - (\eta_t /m) \sum_{i=1}^{m} \nabla_w \psi(w_t, x_i, y_i) \qquad (6)$$

Phương pháp giảm gradient ngẫu nhiên (SGD) thực hiện đơn giản bước cập nhật w_{t+1} chỉ sử dụng một phần tử ngẫu nhiên (x_t, y_t) tại mỗi lần lặp:

$$w_{t+1} = w_t - \eta_t \nabla_w \psi(w_t, x_t, y_t) \qquad (7)$$

Có thể thấy rằng giải thuật SGD đơn giản, thực hiện các bước lặp, mỗi bước lặp chỉ lấy 1 phần tử ngẫu nhiên từ tập dữ liệu, thực hiện cập nhật *w* thay vì phải giải bài toán quy hoạch toàn phương (1). Giải thuật SGD có độ phức tạp tuyến tính với số phần tử của tập dữ liệu học, phân lớp dữ liệu có số phần tử và số chiều lớn rất hiệu quả.

2.3 Thay thế hàm hinge loss bởi hàm xấp xỉ liên tục

Đại lượng lỗi $z_i = max\{0, 1 - y_i(w.x_i)\}$ trong vấn đề tối ưu (5) của máy học SVM thường được gọi là hàm lỗi hinge loss được viết dưới dạng:

$$L^{hinge}(x) = max\{0, 1 - x\} \qquad (8)$$

Chú ý rằng hàm hinge loss không khả vi tại $y_i(w.x_i)=1$. Điều này ảnh hưởng đến tốc độ hội tụ đến lời giải của giải thuật SGD.

Chúng tôi đề xuất thay thế hàm hinge loss bằng các hàm xấp xỉ khả vi để cải tiến tốc độ hội tụ của giải thuật SGD. Một hàm xấp xỉ khả vi của hinge loss, được gọi là smooth hinge loss (Rennie, 2004) có dạng như sau:

$$L^{shinge}(x) = \begin{cases} \frac{1}{2} - x & if\ (x \leq 0) \\ \frac{1}{2}(1-x)^2 & if\ (0 < x < 1) \\ 0 & if\ (x \geq 1) \end{cases} \qquad (9)$$

Ngoài ra, có thể sử dụng hàm xấp xỉ khả vi khác của hinge loss đó là hàm logistic loss (logit), có dạng như sau:

$$L^{logit}(x) = log(1 + e^{-x}) \qquad (10)$$

Hình 2 là đồ thị của hàm hinge loss so với hai hàm xấp xỉ khả vi là smooth hinge loss và logit loss.

Quan sát trên đồ thị, có thể thấy rằng hàm smooth hinge loss và logit loss là hàm khả vi. Nên khi thay thế cho hàm hinge loss có thể đảm bảo được tốc độ hội tụ của giải thuật SGD trong giải trực tiếp vấn đề (5) của máy học SVM. Mặc dù logit loss là hàm trơn (smooth) nhất nhưng smooth hinge loss cũng đủ trơn và vẫn duy trì được tính chất thưa trong lời giải như hàm hinge loss.