$\boxed{\text{Stochastic GD}}$

take $|S| = B$ (mini batch)

$$\frac{1}{B}\sum_{i \in S} L'(y_i f(x_i; w))\, y_i \nabla_w f(x_i; w)$$

$\boxed{\text{SVM I}}$

$\boxed{\text{SVM}}$: ① max margin predictors

② nonlinear SVM (via kernel)

$\boxed{\text{Hard-margin SVM}}$ (training error zero)

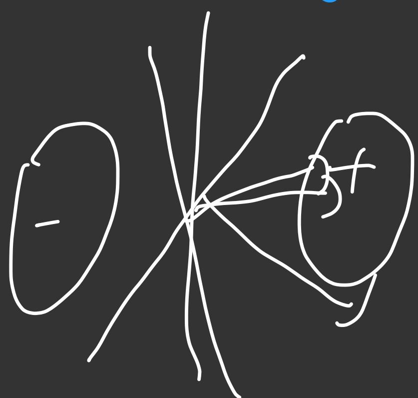
$\boxed{\text{Soft-margin SVM}}$ (training error nonzero)

$\boxed{\text{Nonlinear SVM}}$ kernels

Hard-Margin SVM

- Linear ~ seperable data

$$\min \underbrace{y_i \, x_i^T u}_{\text{sign agree}} > 0 \qquad u \in R^d$$

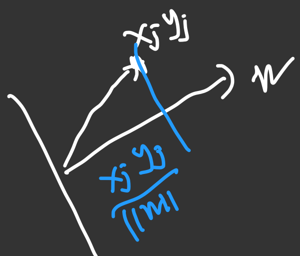


∞ bound, we want the max margin.

$$\min_i w^T x_i y_i > 0$$ ← we should normalize so we can't choose margin by choosing w.

$$\min_i \frac{w^T}{\|w\|} x_i y_i$$ | choose $w^T x_i y_i \geq 1$

$$j = \arg\min y_i x_i^T w$$



$\longrightarrow \frac{1}{\|w\|}$

so
$$\begin{cases} \max \ \frac{1}{\|w\|} \\ \text{subj. } y_i x_i^T w \ge 1 \quad \forall i \end{cases}$$

$$\begin{cases} \min \ \frac{1}{2}\|w\|^2 & \text{(convex)} \\ \text{subj to } y_i x_i^T w \ge 1 \quad \forall i & \text{(convex set)} \end{cases}$$

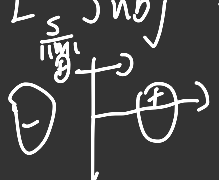
$\boxed{\text{Soft-margin SVM}}$

Suppose the hard margin program

non-feasible.

Introduce slack variables $\xi_1, \dots, \xi_n \ge 0$ $\quad C > 0$ (tradeoff)

$$\begin{cases} \min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i & \text{(convex)} \\ \text{subj to } y_i x_i^T w \ge 1 - \xi_i \quad \forall i & \text{(convex set)} \\ \xi_i \ge 0 \text{ for } i = 1, \cdots, n & \end{cases}$$



Transform previous program to unconstrained

hinge loss

$$\min_{w \in \mathbb{R}^d} \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{n} [1 - y_i x_i^T w]_+$$

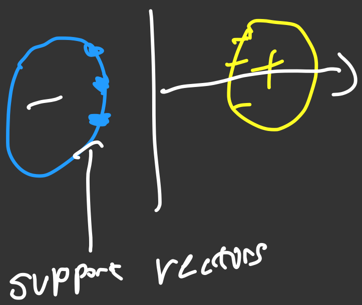
$[a]_+ = \max\{0, a\}$ (RELU)

$$\min_{w \in \mathbb{R}^d} \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{n} L_{hinge}(y_i x_i^T w)$$

$L_{hinge}(z) = \max\{0, 1-z\}$

Each convex program has correlated dual program

Support Vectors



$$L(w,\alpha) = \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i x_i^T w)$$

$$\underbrace{\sup_{\alpha \geq 0} L(w,\alpha)}_{p(w)} = \begin{cases} \frac{1}{2}\|w\|_2^2 & \forall i \; w^T x_i y_i \geq 1 \quad \text{(feasible set)} \\ \infty & \text{(unfeasible case)} \end{cases}$$

$$\Rightarrow \quad \arg\min_{w} p(w) = \arg\min_{w} \sup_{\alpha \geq 0} L(w,\alpha)$$

$$\sup_{\alpha \geq 0} \min_{w} L(w,\alpha)$$

$$\boxed{D(\alpha) = \min_{w} L(w,\alpha) = \sum \alpha_i - \frac{1}{2}\Big\| \sum_i \alpha_i y_i x_i \Big\|^2 \qquad \text{Dual form}}$$