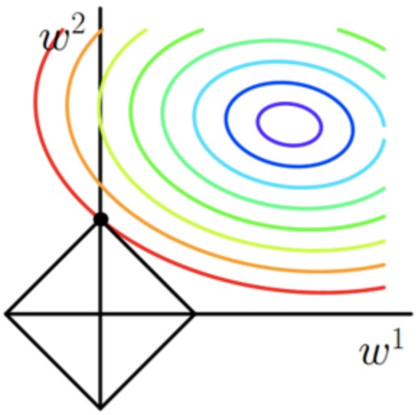
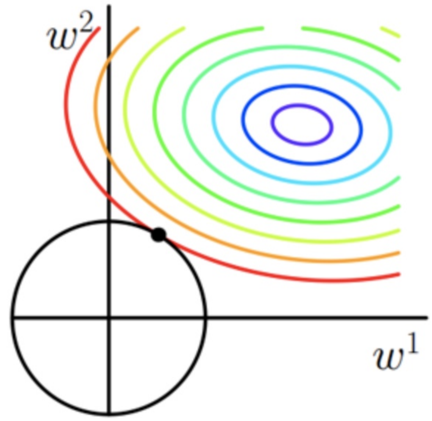
L_1 distance

$|w_1 + w_2|$



L_2 distance

$(w_1^2 + w_2^2)$



L_2 penalty

$\|w\|^2$

If we could solve the optimization problem above, we'd be done. But the "$||w|| = 1$" constraint is a nasty (non-convex) one, and this problem certainly isn't in any format that we can plug into standard optimization software to solve. So, let's try transforming the problem into a nicer one. Consider:

$$\max_{\hat{\gamma},w,b} \quad \frac{\hat{\gamma}}{||w||}$$
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, m$$

$\hat{\gamma}$ geometric margin

Scale constraint

$$\max \frac{1}{\|w\|} \iff \min \|w\|^2$$

$h \to 0$

$$f(x+h) = f(x) + f'(x) \cdot h + \frac{1}{2} f''(x) h^2 + o(h^3)$$

$$f(x-h) = f(x) + f'(x)(-h) + \frac{1}{2} f''(x)(-h)^2 + o(h^3)$$

$$\frac{f(x+h) - f(x-h)}{2h} = \frac{2f'(x) \cdot h}{2h} = f'(x)$$

assignment:

Matrix

$$\begin{array}{cc} & k \downarrow \\ h & \begin{bmatrix} \text{—} \; i \; \text{—} \\ \\ \end{bmatrix} \\ & w \end{array} \quad \begin{array}{cc} \begin{bmatrix} | \\ j \\ | \end{bmatrix} & \leftarrow k \end{array}$$

KNN

two loop

$$\begin{bmatrix} \text{—} \; x_1 \; \text{—} \\ \text{—} \; x_2 \; \text{—} \\ \vdots \\ \text{—} \; x_n \; \text{—} \end{bmatrix} \quad \begin{bmatrix} \text{—} \; y_1 \; \text{—} \\ \text{—} \; y_2 \; \text{—} \\ \vdots \\ \text{—} \; y_N \; \text{—} \end{bmatrix}$$

$$[\text{—} \; x_2 \; \text{—}] - [\text{—} \; y_1 \; \text{—}]$$

one loop

broadcast

$$\begin{bmatrix} \text{—} \; y_1 \; \text{—} \\ \text{—} \; y_2 \; \text{—} \\ \vdots \\ \text{—} \; y_N \; \text{—} \end{bmatrix} - [\text{—} \; x_i \; \text{—}] \Rightarrow \begin{bmatrix} \text{—} \; x_i \; \text{—} \\ \text{—} \; x_i \; \text{—} \\ \vdots \\ \text{—} \; x_i \; \text{—} \end{bmatrix}$$

vector

$$dist = \begin{bmatrix} x_1^2 - 2x_1y_1^T + y_1^2 & x_1^2 - 2x_1y_2^T + y_2^T, & \cdots \\ x_2^2 - 2x_2y_1^T + y_1^2 & x_2^2 - 2x_2y_2^T + y_2^T, & \cdots \\ \vdots & & \end{bmatrix}$$

$$= \begin{bmatrix} x_1^2 & x_1^2 & \cdots \\ x_2^2 & x_2^2 & \cdots \end{bmatrix} + \begin{bmatrix} y_1^2 & y_2^2 & \cdots \\ y_1^2 & y_2^2 & \cdots \\ \vdots & & \end{bmatrix} \rightarrow \begin{bmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_N^2 \end{bmatrix}$$

$$\begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_N^2 \end{bmatrix} \leftarrow$$

$$- 2 \begin{bmatrix} x_1y_1^T & x_1y_2^T & \cdots \\ x_2y_1^T & x_2y_2^T & \cdots \\ \vdots & & \end{bmatrix}$$

$$\begin{bmatrix} - & x_1 & - \\ - & x_2 & - \\ & \vdots & \end{bmatrix} \cdot \begin{bmatrix} | & | & \\ y_1^T & y_2^T & \cdots \\ | & | & \end{bmatrix}$$

SVM

$$\underset{N\times D}{X} \cdot \underset{D\times C}{W} = \underset{N\times C}{S}$$

$$\begin{bmatrix} - & x_1 & - \\ - & x_2 & - \\ & \vdots & \\ - & x_i & - \\ & \vdots & \\ - & x_N & - \end{bmatrix}_{N\times D} \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_C \\ | & | & & | \end{bmatrix}_{D\times C} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1C} \\ S_{21} & S_{22} & \cdots & S_{2C} \\ & \vdots & & \\ S_{N1} & & \cdots & S_{NC} \end{bmatrix}_{N\times C} \begin{matrix} \to L_1 \\ \to L_2 \\ \vdots \\ \to L_N \end{matrix}$$

$$\frac{\partial \text{Loss}}{\partial W} = \frac{1}{N}\frac{\partial(\sum_i L_i)}{\partial W} = \frac{1}{N}\sum_i \frac{\partial L_i}{\partial W}$$

$$L_i = \sum_{j\neq y_i} \max(0, \underbrace{w_j x_i - w_{y_i} x_i + 1}_{L_{ij}}) = \sum_{j\neq y_i} L_{ij}$$

$$\frac{\partial L_{ij}}{\partial W} = \begin{bmatrix} | & | & & | \\ \frac{\partial L_{ij}}{\partial w_1} & \frac{\partial L_{ij}}{\partial w_2} & \cdots & \frac{\partial L_{ij}}{\partial w_C} \\ | & | & & | \end{bmatrix}$$

$$\frac{\partial L_{ij}}{\partial w_k} = 0$$

$$\frac{\partial L_{ij}}{\partial w_j}(\text{score}>0) = x_i \qquad \frac{\partial L_{ij}}{\partial w_{y_i}}(\text{score}>0) = -x_i$$

vector

$$\begin{array}{ccc} X & W & S \end{array}$$

$$\begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_N & - \end{bmatrix}_{N\times D} \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_c \\ | & | & & | \end{bmatrix}_{D\times C} = \begin{bmatrix} - & s_1 & - \\ - & s_2 & - \\ & \vdots & \\ - & s_N & - \end{bmatrix}_{N\times C} \begin{matrix} \to L_1 \\ \to L_2 \\ \\ \to L_N \end{matrix}$$

$$X \cdot W = S \qquad f(s) = l$$

$$\underset{D\times C}{\frac{\partial l}{\partial W}} = \underset{D\times N}{X^T} \underset{N\times C}{\frac{\partial l}{\partial S}} = X^T \begin{bmatrix} - & \frac{\partial l}{\partial s_1} & - \\ - & \frac{\partial l}{\partial s_2} & - \\ & \vdots & \end{bmatrix}$$

$$\frac{\partial l}{\partial s_j} = \sum_i \frac{\partial L_i}{\partial s_j} = \frac{\partial l_j}{\partial s_j} = \left[\frac{\partial l_j}{\partial s_{1j}} \quad \frac{\partial l_j}{\partial s_{2j}} \quad \cdots \quad \frac{\partial l_j}{\partial s_{cj}} \right]$$

$$l_j = \sum_{k \neq y_j} \max(0,\ s_{kj} - s_{y_j j} + \Delta)$$

$$= \max(0,\ s_{1j} - s_{y_j j} + \Delta) + \max(0,\ s_{2j} - s_{y_j j} + \Delta)$$

$$+ \cdots$$

$$\frac{\partial l_j}{\partial s_j} = \left[\frac{\partial l_j}{\partial s_{1j}},\ \frac{\partial l_j}{\partial s_{2j}} \cdots \frac{\partial l_j}{\partial s_{y_j j}} \quad \frac{\partial l_j}{\partial s_{cj}} \right]$$

$$\downarrow$$

$$[0, 1, 1, -3, 1, 0, 0]$$