Lagrangian multipliers:

\[
\min_w f(w) \\
\text{subject to:} \quad h_i(w) = 0 \quad : \beta_i \\
L(w, \beta_1, \ldots, \beta_n) = f(w) + \beta_1 h_1(w) + \ldots + \beta_n h_n(w)
\]

Solution:
\[
\begin{align*}
\frac{\partial L}{\partial w} &= 0 \\
\vdots \\
\frac{\partial L}{\partial \beta_i} &= 0 \\
\vdots
\end{align*}
\]

system of \(n + d\) equations

the dimension of \(w\)

Lagrangian duality, general case, also allowing inequalities:

\[
\min_w f(w) \\
g_i(w) \leq 0 \quad : \alpha_i \\
h_j(w) = 0 \quad : \beta_j \\
L(w, \alpha, \beta) = f(w) + \sum_i \alpha_i g_i(w) + \sum_j \beta_j h_j(w)
\]

Primal problem: \(\min_w E_p(w)\) where \(E_p(w) = \max_{\alpha, \beta} L(w, \alpha, \beta)\)

Dual problem: \(\max_{\alpha, \beta} E_d(\alpha, \beta)\) where \(E_d(\alpha, \beta) = \min_w L(w, \alpha, \beta)\)

Weak duality: \(P^* \geq D^*\) where \(P^* = \min \max L\) and \(D^* = \max \min L\)
Karush - Kuhn - Tucker (KKT) Conditions:

\[ \frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0, \quad \forall i \]

\[ \frac{\partial}{\partial \beta_i} L(w^*, \alpha^*, \beta^*) = 0, \quad \forall i \]

Special case of SVM, solve dual. Find \( E_p(\alpha, \beta) = \min_w L(w, \alpha, \beta) \):

\[ \min \frac{1}{2} w^T w \quad \text{subject to} \quad \|w\| = 1 \]

\[ y_i (w^T x_i + b) \geq 1 \Rightarrow 1 - y_i (w^T x_i + b) \leq 0, \quad \forall i \quad : \alpha_i \]

\[ L(w, \alpha, \beta) = f(w) + \sum_i \alpha_i g_i(w) + \frac{1}{2} \beta^T \beta h(w) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \alpha_i [1 - y_i (w^T x_i + b)] \]

Solution:

\[ \frac{\partial}{\partial w} L(w, \beta, \alpha) = 0 = w + \sum_{i=1}^{N} \alpha_i [y_i x_i] \Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i \]

\[ \frac{\partial}{\partial b} L(w, \beta, \alpha) = \sum_{i=1}^{N} \alpha_i y_i = 0 \]

2) Plug solution for \( w \) into loss, find solution for \( \alpha \):

\[ L(w, b, \alpha) = \frac{1}{2} \left[ \sum_{j=1}^{N} \alpha_j y_j x_j \right]^T \left[ \sum_{i=1}^{N} \alpha_i y_i x_i \right] + \sum_{i=1}^{N} \alpha_i [1 - y_i (\sum_{j=1}^{N} \alpha_j y_j x_j)^T x_i + b)] \]

\[ = \frac{1}{2} \left[ \sum_{j=1}^{N} \alpha_j y_j x_j \right]^T \left[ \sum_{i=1}^{N} \alpha_i y_i x_i \right] + \sum_{i=1}^{N} \alpha_i - \left[ \sum_{j=1}^{N} \alpha_j y_j x_j \right]^T \left[ \sum_{i=1}^{N} \alpha_i y_i x_i \right] - \sum_{i=1}^{N} \alpha_i 0 \]
\[ (3) \quad = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \left( \sum_{i=1}^{N} \alpha_i y_i x_i \right)^T \left( \sum_{j=1}^{N} \alpha_j y_j x_j \right) \]

Hence solve SVM dual as QP:

\[
\max_{\alpha} \quad \alpha^T 1 - \frac{1}{2} \alpha^T H \alpha \quad \text{where } H_{ij} = y_i y_j x_i^T x_j \\
\text{s.t. } \sum_{i=1}^{N} \alpha_i y_i = 0 \\
\alpha_i \geq 0, \quad y_i
\]

QP is concave because \( H \) is positive semi-definite:

\( H = A^T A \) (where \( A \) is a matrix with columns \( y_j x_j \)) so

\( \alpha^T H \alpha = \alpha^T A^T A \alpha = \| A \alpha \|^2 \geq 0 \), \( \alpha \) convex

Why do we care about dual?

#1 It exposes structure: a) \( W = \sum_{i=1}^{N} \alpha_i y_i x_i \)

b) KKT#3 \( \alpha_i y_i (w) = 0 \rightarrow \text{if } \alpha_i \geq 0 \text{ then } g_i(w) = 0 \)

i.e. if \( \alpha_i \geq 0 \) then \( 1 - y_i (w^T x_i + b) = 0 \rightarrow y_i (w^T x_i + b) = 1 \)

i.e. Simple \( i \) is a SV and lies on the margin!

so \( W = \sum \alpha_i y_i x_i \) and non-SV points don't matter i.e., SV, \( SV \) is the set of support vectors

#2 The kernel trick: data only appears in pairs as \( x_i^T x_j \):

\( \hat{W}^T x + b = \sum_{i=1}^{N} \alpha_i y_i \frac{x_i^T x_j}{x_i^T x_i} + b \) (prediction for \( x \) before threshold)