Proposition: (Maximum Margin Classifier) Given a linearly separable training set

\[ D = \{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \ldots, (\bar{x}_l, y_l)\} \subseteq \mathbb{R}^n \times \{+1, -1\}, \]

we can compute a maximum margin decision surface \( \overline{w}^* \cdot \bar{x} = b^* \) with an optimization,

\[
\min_{\overline{w}, b} \phi(\overline{w}, b) = \min_{\overline{w}, b} \frac{1}{2} \overline{w} \cdot \overline{w}
\]

subject to the constraints,

\[ \overline{w} \cdot (y_i \bar{x}_i) \geq 1 + y_i b \quad \text{for all} \ (\bar{x}_i, y_i) \in D. \]
Maximum Margin Classifiers

\[ \overline{w}^* \cdot \overline{x} = b^* + k \]

\[ \overline{w}^* \cdot \overline{x} = b^* \]

\[ \overline{w}^* \cdot \overline{x} = b^* - k \]

\[ m^* \]
Our objective function is convex,

$$\phi(w, b) = \frac{1}{2} \bar{w} \cdot \bar{w} = \frac{1}{2} (w_1^2 + \ldots + w_n^2),$$

Here $\bar{w} \in \mathbb{R}^2$. 
A quadratic program is a general convex optimization problem of the form

$$\overline{w}^* = \arg\min_{\overline{w}} \left( \frac{1}{2} \overline{w}^T Q \overline{w} - \overline{q} \cdot \overline{w} \right),$$

subject to the constraints

$$X^T \overline{w} \geq \overline{c}.$$

Here, $Q$ is an $n \times n$ matrix, $X$ is an $l \times n$ matrix, the vectors $\overline{w}^*$, $\overline{w}$, $\overline{q}$ are $n$-dimensional vectors, and the vector $\overline{c}$ is an $l$-dimensional vector. 

In software packages this is usually given as function of the form,

$$\overline{w}^* = \text{solve}(Q, \overline{q}, X, \overline{c}).$$

---

*I have written the quadratic program in terms of $\overline{w}$; in the literature a different letter would typically be used for the optimization variable.*
In order to bring the generalized quadratic program into a form that we can use for our maximum margin optimization we let

\[ Q = I, \]

and

\[ \bar{q} = 0, \]

then

\[ \bar{w}^* = \arg \min_{\bar{w}} \left( \frac{1}{2} \bar{w}^T I \bar{w} - \bar{0} \cdot \bar{w} \right) = \arg \min_{\bar{w}} \left( \frac{1}{2} \bar{w} \cdot \bar{w} \right), \]
Next, let us look at the original constraints,

$$(y_i \overline{x}_i) \cdot \overline{w} \geq 1 + y_i b,$$

for all $$(\overline{x}_i, y_i) \in D$$ with $$i = 1, \ldots, l$$ and $$\overline{x}_i = (x^1_i, \ldots, x^n_i)$$.

We have to rewrite these into the matrix form,

$$X^T \overline{w} \geq \overline{c},$$

with

$$X = \begin{pmatrix} y_1 x^1_1 & \cdots & y_1 x^1_i & \cdots & y_l x^1_i \\
\vdots & & \vdots & & \vdots \\
y_1 x^n_1 & \cdots & y_1 x^n_i & \cdots & y_l x^n_i \end{pmatrix} \quad \overline{c} = \begin{pmatrix} 1 + y_1 b \\
1 + y_2 b \\
\vdots \\
1 + y_l b \end{pmatrix}$$

**Observation:** $b$ is now a free variable in the optimization problem, its value is not computed by the optimization algorithm but must be set by the user.
Proposition: (Quadratic Programming) Given a linearly separable training set

\[ D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_l, y_l)\} \subseteq \mathbb{R}^n \times \{+1, -1\}, \]

then we can compute a maximum margin decision surface \( \overline{w}^* \cdot \mathbf{x} = b^* \) with a quadratic programming approach that solves the generalized optimization problem,

\[
(\overline{w}^*, b^*) = \operatorname{argmin}_{\overline{w}, b} \left( \frac{1}{2} \overline{w}^T Q \overline{w} - \overline{q} \cdot \overline{w} \right),
\]

subject to the constraints

\[ \mathbf{X}^T \overline{w} \geq \overline{c}, \]

with \( Q = \mathbf{I}, \overline{q} = \overline{0} \), and where \( \mathbf{X} \), and \( \overline{c} \) are constructed according to the previous discussion.
let $D = \{(\overline{x}_1, y_1), (\overline{x}_2, y_2), \ldots, (\overline{x}_l, y_l)\} \subset \mathbb{R}^n \times \{+1, -1\}$

$r \leftarrow \max\{|\overline{x}| \mid (\overline{x}, y) \in D\}$

$q \leftarrow 1000$

let $\overline{w}^*$ and $b^*$ be undefined

construct $X$

for each $b \in [-q, q]$ do

construct $\overline{c}$

$\overline{w} \leftarrow \text{solve}(I, \overline{0}, X, \overline{c})$

if $(\overline{w}$ is defined and $\overline{w}^*$ is undefined) or

$(\overline{w}$ is defined and $|\overline{w}| < |\overline{w}|^*)$ then

$\overline{w}^* \leftarrow \overline{w}$

$b^* \leftarrow b$

end if

end for

if $\overline{w}^*$ is undefined then stop constraints not satisfiable

else if $|\overline{w}|^* > q/r$ then stop bounding assumption of $|\overline{w}|$ violated

end if

return $(\overline{w}^*, b^*)$
Let

\[ D = \{((1, 6), -1), ((3, 7), -1), ((1, 4), +1), ((2, 1), +1)\} \]

be the training set and let

\[ \overline{w^*} = solve(Q, \overline{q}, X, \overline{c}), \]

be a call to the solver, then

\[ X = \begin{bmatrix} -1 & -3 & 1 & 2 \\ -6 & -7 & 4 & 1 \end{bmatrix} \]

\[ \overline{c} = \begin{bmatrix} 1 - b \\ 1 - b \\ 1 + b \\ 1 + b \end{bmatrix} \]

\[ Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ \overline{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]
**Observation:** Our optimization problem has a free parameter, the offset term $b$. Notice also that the constraints are dependent on this term. That is we need to pick $b$ in such a way that the constraints are consistent. In other words, we need to pick $b$ in such a way that the quadratic solver can actually find a minimum $\overline{w}$.

What are reasonable values for $b$ to try?
First observation,

\[ b = \overline{w} \cdot \overline{x} \quad \Rightarrow \]

\[ b = |\overline{w}| |\overline{x}| \cos \gamma \quad \Rightarrow \quad (-1 \leq \cos \gamma \leq 1, \text{for all } \gamma) \]

\[ -|\overline{w}| |\overline{x}| \leq b \leq |\overline{w}| |\overline{x}| \quad \Rightarrow \quad (|\overline{x}| \leq r) \]

\[ -|\overline{w}| r \leq b \leq |\overline{w}| r \quad \Rightarrow \]

Second observation, \( 2r \) is the size of the largest margin and \( \overline{w} \) is unbounded,

\[ \frac{2}{|\overline{w}|} \leq 2r \Rightarrow \frac{1}{r} \leq |\overline{w}| \]

Third observation, we bound \( \overline{w} \),

\[ \frac{1}{r} \leq |\overline{w}| \leq \frac{q}{r} \]

Finally, we use \( \overline{w} = \frac{q}{r} \) in the equation above,

\[ -|\overline{w}| r \leq b \leq |\overline{w}| r \Rightarrow -q \leq b \leq q \]