Instead of developing SVMs via Langrangian optimization theory we can develop SVM using convex hulls.
Convex Hulls

Let \( X = \{ \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_l \} \subset \mathbb{R}^n \), then the convex hull of \( X \) is the set of all convex combinations of its points, \( H(X) \). In a convex combination, each point in is assigned a weight or coefficient in such a way that the coefficients are all non-negative and sum to one, and these weights are used to compute a weighted average of the points. For each choice of coefficients, the resulting convex combination is a point in the convex hull, and the whole convex hull can be formed by choosing coefficients in all possible ways. Expressing this as a single formula, the convex hull is the set:

\[
H(X) = \left\{ \sum_{i=1}^{l} \alpha_i \bar{x}_i \right\}
\]

with \( \sum_{i=1}^{l} \alpha_i = 1 \) and \( \alpha_i \geq 0 \).
SVM: The Separable Case

Let \( D \) be our training data. Consider two class distributions +1 and -1 and their corresponding hulls \( H(+1) \) and \( H(-1) \),

We pick the point \( \bar{c} \in H(+1) \) to be closest to the -1 class distribution and we pick point \( \bar{d} \in H(-1) \) to be closest to the +1 distribution. Next we draw a vector from \( \bar{d} \) to \( \bar{c} \) such that

\[
\bar{w} = \bar{c} - \bar{d}
\]

Now, picking the points \( \bar{c} \) and \( \bar{d} \) as we did above and then drawing the vector \( \bar{w} \) is the same as saying that we want to minimize the length of \( \bar{w} \), in other words,

\[
\min |\bar{w}| = \min \frac{1}{2} |\bar{w}|^2 = \min \frac{1}{2} \bar{w} \cdot \bar{w}
\]
SVM: The Separable Case

Now consider that \( c \in H(+1) \) and \( d \in H(-1) \), therefore

\[
\begin{align*}
\overline{c} &= \sum_{x_p \in +1} \alpha^+_p x_p \\
\overline{d} &= \sum_{x_q \in -1} \alpha^-_q x_q
\end{align*}
\]

Now, let \( \overline{\alpha} \) be the \textit{concatenation} of \( \overline{\alpha}^+ \) and \( \overline{\alpha}^- \) with

\[
|\overline{\alpha}| = |\overline{\alpha}^+| + |\overline{\alpha}^-| = l
\]

then

\[
\min_{\alpha} \frac{1}{2} \overline{w} \cdot \overline{w} = \min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j x_i \cdot x_j
\]

subject to

\[
\sum_{i=1}^{l} y_i \alpha_i = 0
\]

\[
\alpha_i \geq 0
\]
SVM: The Separable Case

It is worthwhile to take a look at the constraint

$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

We can rewrite this constraint as

$$\sum_{i=1}^{\alpha^+} (+1) \alpha_i^+ + \sum_{i=1}^{\alpha^-} (-1) \alpha_i^- = \sum_{i=1}^{\alpha^+} \alpha_i^+ - \sum_{i=1}^{\alpha^-} \alpha_i^- = 1 - 1 = 0$$ (if the points fulfill the convex hull property)
Finally, we have

\[
\max_\alpha -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \bar{x}_i \cdot \bar{x}_j
\]

subject to

\[
\sum_{i=1}^{l} y_i \alpha_i = 0
\]
\[
\alpha_i \geq 0
\]

It is interesting to note that this looks very similar to the optimization problem that we derived via Lagrangian optimization theory.
The reduced hull $RH(X)$ is defined as

$$RH(X) = \left\{ \sum_{i=1}^{l} \alpha_i x_i \right\}$$

with

$$\sum_{i=1}^{l} \alpha_i = 1$$

and

$$C \geq \alpha_i \geq 0$$
SVM: The Non-Separable Case

The optimization problem then becomes

\[
\max_{\alpha} - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j
\]

subject to

\[
\sum_{i=1}^{l} y_i \alpha_i = 0
\]

\[
C \geq \alpha_i \geq 0
\]
Model

It is easy to show that our model is a support vector machine,

\[ \hat{f}(x) = \text{sign}(\overline{w}^* \cdot x - b^*) \]

with

\[ \overline{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \text{ (think } \overline{w} = \overline{c} - \overline{d}) \]

and

\[ b^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \cdot \overline{x}_{sv} + 1 \]
C. Bennet – "SVM - Hype or Hallelujah" – available on the course website.