SVM Formulation of the Portfolio Optimization Problem

Thierry Bazier-Matte

December 29, 2016

1 Linearly Separable and Step Utility Function

Under an SVM formulation of our problem, we look for a decision q such that $q^T x$ is the same sign as its output. Under a classical SVM formulation, where the problem is one of classification, corresponding to a loss step function, the output is $y = \pm 1$.

Let us suppose for now that returns r_i are linearly separable. Then we can scale the decision q to obtain $\min_S |q^T x| = 1$, with S the sample. Then, the distance of any feature point $x_0 \in \mathscr{R}^p$ will be given by

$$\frac{|q^T x_0|}{\|q\|_2}.$$

Therefore the margin ρ of the SVM will be given by

$$\rho = \frac{1}{\|q\|_2}.$$

Geometrically, we wih to have a margin as large as possible or, equivalently, to minimize $\|q\|_2$, or more simply $\frac{1}{2}\|q\|_2^2$.

Under the classical formulation of the SVM, where $y_i = \pm 1$, we wish to have $\operatorname{sign} q^T x_i = \operatorname{sign} y_i$. This can also be expressed as

$$y_i \, q^T x_i \geq 1,$$

since $q^T x_i \geq 1$. Under the porfolio formulation, the output y_i is actually the portfolio return $r_i \in \mathscr{R}$. Let us pretend for the moment that we are endowed with a step utility function. In other words, we only care for the sign of our investment. Then we would again like to obtain $\operatorname{sign} r_i = \operatorname{sign} q^T x_i$. Let us further suppose that there is no j such that $r_j = 0$. Let $\bar{r} = \min_S r_i$. Then using the transformation $\tilde{r}_i = r_i/\bar{r}$, we obtain $r_i \geq 1$. We therefore obtain the same formulation as with the SVM, that is,

$$\begin{array}{ll} \text{minimize} & \|q\|^2 \\ \text{subject to} & \tilde{r}_i\, q^T x_i \geq 1. \end{array}$$

2 Utility dependant case

Under the classical SVM formulation, when the set is not separable, we can add slack variables $\xi_i \geq 0$ to the margin so that

$$y_i\, q^T x_i \geq 1 - \xi_i.$$

The SVM formulation can therefore be expressed as

$$\begin{array}{ll} \text{minimize} & \|q\|^2 + \lambda\|\xi\|^p \\ \text{subject to} & y_i\, q^T x_i \geq 1 - \xi_i \\ & \xi_i \geq 0, \end{array}$$

where μ is the trade off parameter between the slack and the margin width.

Under the classical framework, ξ_i is only some slack, and it loss function will typically be quadratic or linear (hinge loss). However, under the portfolio optimization framework, this slack ξ_i actually means much more since we can encourage or discourage going in certain directions proportionally to the utility of the investor.

Let us first suppose that we have no preference for positive returns, or simply that we have a flat utility for $r \geq 0$. Such an asumption corresponds to $\xi_i \geq 0$: as long as the returns $r_i\, q^T x_i$ are deep enough out of the margin, we are satisfied. But we can now quantify how much dissatisfaction we retire from going toward the margin. We can impose a penalty $-u(\xi_i)$ on the slack on the whole domain. We can further drop the absence of preference on positive returns, by simply stating that we impose $-u(\xi_i)$ on the slack. This way, the utility perception of the investor is transposed in the SVM algorithm using the slacks:

$$\begin{array}{ll} \text{minimize} & \|q\|^2 - \lambda \sum_{i=1}^{n} u(\xi_i) \\ \text{subject to} & \tilde{r}_i\, q^T x_i \geq 1 - \xi_i \end{array}$$

The optimization problem remains convex, since u is concave.

Under such a formulation, we see how close have gotten to the original problem: compare this last expression with the following:

$$\text{maximize} \quad \frac{1}{n} \sum_{i=1}^{n} u(r_i\, q^T x_i) - \lambda\|q\|^2.$$

It is pratically the same: we now see how close to an actual SVM we are. In wit, this means that are theoretical properties of SVMs also apply to our problem. [**Todo: Show formally.**]

Now what interpretation can be given to those two traded off terms? The first expression $\|q\|^2$ means how wide the margin will be, that is how probable it is that we actually have overall positive returns. It can also be meant to represent the complexity of the solution: a tighter margin will accomodate more outlier points, thus reducing their numbers and thus favouring a more complex solution, whereas a wider margin will be simpler, albeit with more outliers. The other expression $-\sum u(\xi_i)$ can be understood as the average of the utility for each observed return $r_i\, q^T x_i$.