RJ10524 (ALM1410-003) October 2, 2014
Computer Science

IBM Research Report

SVM Computation in a Distributed Database

Nimrod Megiddo
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099
USA

SVM Computation in a Distributed Database

Nimrod Megiddo

July 25, 2013

1 Introduction

The so-called support vector machines (SVM) have become a very important tool for the classification problem. Computing an SVM amounts to solving a certain optimization problem. The SVM optimization problem is posed with respect to a set of labeled examples given explicitly. In real-life databases, the data is often distributed over various tables. Even if the data is given in a single table, there are often external sources of data that can improve the accuracy of a classifier if incorporated in the classifier. For example, a given table providing attributes of individuals that have to be classified may include the town where the individual resides but no attributes of that town. An external source may provide various attributes of towns or transaction that took place in various towns, which may be relevant to the classification of individuals. Thus, it is desirable to build a classifier that takes some of these attributes or transactions into account. This hypothesis calls for joining the tables on the town column.

To apply a standard SVM algorithm when attributes are distributed over tables, one has to first to join the tables. However, joining tables explicitly may not be possible due to the size of the product. Thus, the question is whether it is possible to obtain an SVM for the join without generating the table explicitly. Here, we show how this can be done for the join of two tables. In general, the size of the join of two tables can be quadratic in the terms of the sizes of the joined tables.

1.1 The standard SVM

We first review the standard SVM problem. The input table consists of m "examples" given as feature vectors $\boldsymbol{x}_i \in \Re^d$ and corresponding class labels $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$.

The primal problem

The primal SVM optimization problem is the following:

$$\begin{array}{ll} \text{Minimize}_{\boldsymbol{w},b,\boldsymbol{\xi}} & \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{m} \xi_i \\ \text{subject to} & y_i \boldsymbol{x}_i^\top \boldsymbol{w} - y_i b + \xi_i \geq 1 \quad (i = 1, \ldots, m) \\ & \xi_i \geq 0 \quad (i = 1, \ldots, m). \end{array} \tag{1}$$

The dual problem

The Lagrangian function of the problem in (1) is the following:

$$
\begin{aligned}
L(\boldsymbol{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha}) &= \tfrac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i \left(y_i \boldsymbol{x}_i^\top \boldsymbol{w} - y_i b + \xi_i - 1 \right) \\
&= \tfrac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i^\top \boldsymbol{w} + b \sum_{i=1}^{m} y_i \alpha_i + \sum_{i=1}^{m} \xi_i (C - \alpha_i) + \sum_{i=1}^{m} \alpha_i \ .
\end{aligned} \tag{2}
$$

In the following problem, an optimal solution must satisfy the constraints of (1) and also $\alpha_i = 0$ for every i such that $y_i \boldsymbol{x}_i^\top \boldsymbol{w} - y_i b + \xi_i > 1$:

$$
\text{Minimize}_{\boldsymbol{w}, b, \boldsymbol{\xi}} \left\{ \max_{\boldsymbol{\alpha}} \{ L(\boldsymbol{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha}) : \boldsymbol{\alpha} \geq \mathbf{0} \} \ : \ \boldsymbol{\xi} \geq \mathbf{0} \right\} \ . \tag{3}
$$

It follows that (3) is equivalent to (1). Due to the convexity in terms of $(\boldsymbol{w}, b, \boldsymbol{\xi})$ and linearity in terms of $\boldsymbol{\alpha}$, the optimal value of (3) is equal to the optimal value of the following:

$$
\text{Maximize}_{\boldsymbol{\alpha}} \left\{ \min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \{ L(\boldsymbol{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha}) : \boldsymbol{\xi} \geq \mathbf{0} \} \ : \ \boldsymbol{\alpha} \geq \mathbf{0} \right\} \ . \tag{4}
$$

Let $\boldsymbol{\alpha} \geq \mathbf{0}$ be fixed for a moment. If $\sum_{i=1}^{m} y_i \alpha_i \neq 0$, then $b \sum_{i=1}^{m} y_i \alpha_i$ is not bounded from below. Similarly, if $\alpha_i > C$, then $\xi_i (C - \alpha_i)$ is not bounded from below when $\xi_i > 0$. Therefore, an optimal $\boldsymbol{\alpha}$ for (4) must satisfy

$$
\sum_{i=1}^{m} \alpha_i y_i = 0 \quad \text{and} \quad \alpha_i \leq C \ \ (i = 1, \ldots, m) \ .
$$

Next, the unique $\boldsymbol{w}$ that minimizes $L(\boldsymbol{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha})$ is

$$
\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \ . \tag{5}
$$

Finally, if $\boldsymbol{\xi} \geq \mathbf{0}$ minimizes $L(\boldsymbol{w}, b, \boldsymbol{\xi}; \boldsymbol{\alpha})$, then for every i such that $\alpha_i < C$, necessarily $\xi_i = 0$, and hence

$$
\sum_{i=1}^{m} \xi_i (C - \alpha_i) = 0 \ . \tag{6}
$$

Thus, the problem in(4) is equivalent to the following, which can be viewed as the dual problem:

$$
\begin{aligned}
&\text{Minimize}_{\boldsymbol{\alpha}} \ \tfrac{1}{2} \sum_{ij} y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j \alpha_i \alpha_j - \sum_{i} \alpha_i \\
&\text{subject to} \ \ \sum_{i=1}^{m} y_i \alpha_i = 0 \\
&\qquad\qquad\ \ 0 \leq \alpha_i \leq C \ .
\end{aligned} \tag{7}
$$

2 SVM on a join of two tables

2.1 Formulation

We now consider a problem with *two* tables, T_1 and T_2. The table T_1 has m rows $(\boldsymbol{p}_i^\top, \boldsymbol{u}_i^\top)$, $i = 1, \ldots, m$, and the table T_2 has n rows $(\boldsymbol{q}_j^\top, \boldsymbol{v}_j^\top)$, $j = 1, \ldots, n$, with columns as follows. The attributes that are represented by the columns of these tables are of three types described below. Denote by P the set of attributes represented by the $\boldsymbol{p}_i$s, and by Q the set of attributes represented by the $\boldsymbol{q}_j$s. The set U of attributes represented by the $\boldsymbol{u}_i$s is the same as the set V of attributes represented by the $\boldsymbol{v}_j$s (these are the common attributes of the two tables). The class labels y_i are associated with the rows of T_1. The (universal) *join* of T_1 and T_2 is a new table J, consisting of $|P|+|U|+|Q|$ columns, defined as follows. For each i, $i = 1, \ldots, m$, if there is no j such that $\boldsymbol{u}_i^\top = \boldsymbol{v}_j^\top$, then J has a row $\boldsymbol{x}_{i0}^\top = (\boldsymbol{p}_i^\top, \boldsymbol{u}_i^\top, \boldsymbol{0}^\top)$; otherwise, J has rows of the form $\boldsymbol{x}_{ij}^\top = (\boldsymbol{p}_i^\top, \boldsymbol{u}_i^\top, \boldsymbol{q}_j^\top)$ for every pair (i, j) such that $\boldsymbol{u}_i^\top = \boldsymbol{v}_j^\top$. Denote by $\boldsymbol{w}_P$, $\boldsymbol{w}_U$ and $\boldsymbol{w}_Q$ the projections of the (unknown) vector $\boldsymbol{w}$ on the sets P, U and Q, respectively. Also, denote

$$\mathrm{I}_0 = \{(i, 0) : (\forall j)(\boldsymbol{u}_i \neq \boldsymbol{v}_j)\}$$

and

$$\mathrm{IJ} = \mathrm{I}_0 \cup \{(i, j) : \boldsymbol{u}_i = \boldsymbol{v}_j\} \ .$$

Thus, the explicit form of the primal problem over the join is:

$$\begin{aligned} \text{Minimize}_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad & \tfrac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{(i,j)\in\mathrm{IJ}} \xi_{ij} \\ \text{subject to} \quad & y_i \boldsymbol{x}_{ij}^\top \boldsymbol{w} - y_i b + \xi_{ij} \geq 1 \qquad ((i,j) \in \mathrm{IJ}) \\ & \xi_{ij} \geq 0 \qquad ((i,j) \in \mathrm{IJ}). \end{aligned} \tag{8}$$

The size of the latter may be too large, depending on the size of the set IJ. Our goal is to solve the SVM problem on J without explicitly generating all the rows of J. We can reformulate this problem by first observing that

$$\boldsymbol{x}_{ij}^\top \boldsymbol{w} = \boldsymbol{p}_i^\top \boldsymbol{w}_P + \boldsymbol{u}_i^\top \boldsymbol{w}_U + \boldsymbol{q}_j^\top \boldsymbol{w}_Q \tag{9}$$

where, for convenience, we denote $\boldsymbol{q}_0 = \boldsymbol{0}$.

As a first step, we reduce the number of penalty variables as follows. Instead of using a penalty variable ξ_{ij} for each $(i, j) \in \mathrm{IJ}$, we generate those penalties in the form

$$\xi_{ij} = \eta_i + \zeta_j \tag{10}$$

which makes sense in view of (9) because in an optimal solution

$$\xi_{ij} = \max\{0, 1 - y_i \boldsymbol{x}_i^\top \boldsymbol{w} + y_i b\} \ . \tag{11}$$

Thus, we obtain the following modified optimization problem:

$$\begin{aligned} \text{Minimize}_{\boldsymbol{w},b,\boldsymbol{\eta},\boldsymbol{\zeta}} \quad & \tfrac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{m} J(i) \cdot \eta_i + C \cdot \sum_{j=1}^{n} I(j) \cdot \zeta_j \\ \text{subject to} \quad & y_i \boldsymbol{x}_{ij}^\top \boldsymbol{w} - y_i b + \eta_i + \zeta_j \geq 1 \qquad ((i,j) \in \mathrm{IJ}) \\ & \eta_i, \zeta_j \geq 0 \ , \end{aligned} \tag{12}$$

where $J(i) = |\{j : (i,j) \in \mathrm{IJ}\}|$ and $I(j) = |\{i : (i,j) \in \mathrm{IJ}\}|$.

Note that the number of constraints in problem (12) may still be too large for solving the problem in practice (depending on the size of IJ), so we need to simplify the problem further.

2.2 A linear-size formulation

Denote by $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell$ all the distinct values that appear as $\boldsymbol{u}_i$. For each k, $k = 1, \ldots, \ell$, denote

$$I_k = \{i : \boldsymbol{u}_i = \boldsymbol{z}_k\}$$

and

$$J_k = \{j : \boldsymbol{v}_i = \boldsymbol{z}_k\} \ .$$

Some sets J_k may be empty. Note that the sets $I_1, \ldots, I_\ell$ partition the set $\{1, \ldots, m\}$ and also the sets $J_1, \ldots, J_\ell$ are pairwise disjoint. We introduce auxiliary variables $\sigma_1, \ldots, \sigma_\ell$ and τ_k for $k = 1, \ldots \ell$ such that $J_k \neq \emptyset$. Consider the following system of constraints:

$$\begin{aligned} y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P - y_i b + \eta_i &\geq \sigma_k \quad (i \in I_k,\ k = 1, \ldots \ell) \\ \boldsymbol{q}_j^\top \boldsymbol{w}_Q + \zeta_j &\geq \tau_k \quad (j \in J_k,\ k = 1, \ldots \ell) \\ \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U + \tau_k &\geq 1 \quad (\text{for } k = 1, \ldots \ell \text{ such that } J_k \neq \emptyset) \\ \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U &\geq 1 \quad (\text{for } k = 1, \ldots \ell \text{ such that } J_k = \emptyset) \ . \end{aligned} \tag{13}$$

Proposition 2.1 *A vector* $\boldsymbol{w}$ *satisfies the system*

$$y_i \boldsymbol{x}_{ij}^\top \boldsymbol{w} - y_i b + \eta_i + \zeta_j \geq 1 \qquad ((i,j) \in \mathrm{IJ}) \tag{14}$$

if and only if there exist $\sigma_1, \ldots, \sigma_\ell$ *and* $\tau_1, \ldots, \tau_\ell$ *that together with* $\boldsymbol{w}$ *satisfy the system (13).*

Thus, we obtain the following compact form:

$$\begin{array}{ll} \text{Minimize}_{\boldsymbol{w},b,\boldsymbol{\eta},\boldsymbol{\zeta},\boldsymbol{\sigma},\boldsymbol{\tau}} & \frac{1}{2}\|\boldsymbol{w}_P\|^2 + \frac{1}{2}\|\boldsymbol{w}_U\|^2 + \frac{1}{2}\|\boldsymbol{w}_Q\|^2 + C \cdot \sum_{i=1}^m J(i) \cdot \eta_i + C \cdot \sum_{i=1}^m I(j) \cdot \zeta_i \\ \text{subject to} & y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P - y_i b + \xi_i - \sigma_k \geq 0 \qquad (i \in I_k,\ k = 1, \ldots \ell) \\ & \boldsymbol{q}_j^\top \boldsymbol{w}_Q - \tau_k \geq 0 \qquad (j \in J_k,\ k = 1, \ldots \ell) \\ & \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U + \tau_k \geq 1 \qquad (\text{for } k = 1, \ldots \ell \text{ such that } J_k \neq \emptyset) \\ & \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U \geq 1 \qquad (\text{for } k = 1, \ldots \ell \text{ such that } J_k = \emptyset) \\ & \xi_i \geq 0 \quad (i = 1, \ldots, m). \end{array} \tag{15}$$

At an optimal solution,

$$\sigma_k = \min_{i \in I_k} \left\{ y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P - y_i b + \eta_i \right\}$$

and

$$\tau_k = \min_{j \in J_k} \left\{ \boldsymbol{q}_j^\top \boldsymbol{w}_Q + \zeta_j \right\} \ .$$

The Lagrangian function of the latter is derived as follows. Let $\alpha_i \geq 0$ be multipliers associated with the constraints:

$$y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P - y_i b + \eta_i - \sigma_k \geq 0 \qquad (i \in I_k,\ k = 1, \ldots \ell) \tag{16}$$

and recall that the I_ks are pairwise disjoint. Let $\beta_j \geq 0$ be multipliers associated with the constraints:

$$\boldsymbol{q}_j^\top \boldsymbol{w}_Q + \zeta_j - \tau_k \geq 0 \quad (j \in J_k,\ k = 1, \ldots \ell)\ , \tag{17}$$

and let $\gamma_k \geq 0$ be multipliers associated with the constraints

$$\begin{array}{ll} \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U + \tau_k \geq 1 & (\text{for } k = 1, \ldots \ell \text{ such that } J_k \neq \emptyset) \\ \sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U \geq 1 & (\text{ for } k = 1, \ldots \ell \text{ such that } J_k = \emptyset)\ . \end{array} \tag{18}$$

The Lagrangian function is:

$$\begin{aligned} L(\boldsymbol{w}_P, \boldsymbol{w}_U, \boldsymbol{w}_Q, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\sigma}, \boldsymbol{\tau}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = {} & \tfrac{1}{2}\|\boldsymbol{w}_P\|^2 + \tfrac{1}{2}\|\boldsymbol{w}_U\|^2 + \tfrac{1}{2}\|\boldsymbol{w}_Q\|^2 \\ & + C \cdot \sum_{i=1}^{m} J(i)\eta_i + C \cdot \sum_{j=1}^{n} I(j)\zeta_j \\ & - \sum_{k=1}^{\ell} \sum_{i \in I_k} \alpha_i \left(y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P - y_i b + \eta_i - \sigma_k \right) - \sum_{k=1}^{\ell} \sum_{j \in J_k} \beta_j \left(\boldsymbol{q}_j^\top \boldsymbol{w}_Q + \zeta_j - \tau_k \right) \\ & - \sum_{k: J_k \neq \emptyset} \gamma_k \left(\sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U + \tau_k - 1 \right) - \sum_{k: J_k = \emptyset} \gamma_k \left(\sigma_k + \boldsymbol{z}_k^\top \boldsymbol{w}_U - 1 \right) \end{aligned} \tag{19}$$

Rearranging terms, we obtain

$$\begin{aligned} L(\boldsymbol{w}_P, \boldsymbol{w}_U, \boldsymbol{w}_Q, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\sigma}, \boldsymbol{\tau}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = {} & (\tfrac{1}{2}\|\boldsymbol{w}_P\|^2 - \sum_i \alpha_i y_i \boldsymbol{p}_i^\top \boldsymbol{w}_P) \\ & + (\tfrac{1}{2}\|\boldsymbol{w}_U\|^2 - \sum_k \gamma_k \boldsymbol{z}_k^\top \boldsymbol{w}_U) + (\tfrac{1}{2}\|\boldsymbol{w}_Q\|^2 - \sum_j \beta_j \boldsymbol{q}_j^\top \boldsymbol{w}_Q) \\ & + \sum_{k=1}^{\ell} \gamma_k - b \sum_i y_i \alpha_i \\ & + \sum_i \eta_i (C\, J(i) - \alpha_i) + \sum_j \zeta_j (C\, I(j) - \beta_j) \\ & + \sum_{k=1}^{\ell} \sigma_k \left(\sum_{i \in I_k} \alpha_i - \gamma_k \right) + \sum_{J_k \neq \emptyset}^{\ell} \tau_k \left(\sum_{j \in J_k} \beta_j - \gamma_k \right)\ . \end{aligned} \tag{20}$$

The dual problem is:

$$\text{Maximize}_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}} \left\{ \min{}_{\boldsymbol{w}, b, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\sigma}, \boldsymbol{\tau}} \{ L(\boldsymbol{w}, b, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\sigma}, \boldsymbol{\tau}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) : \boldsymbol{\xi} \geq \boldsymbol{0} \} \ :\ \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \geq \boldsymbol{0} \right\}\ . \tag{21}$$

Let $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ be fixed for the moment. We must have

$$\boldsymbol{w}_P = \sum_i \alpha_i y_i \boldsymbol{p}_i \tag{22}$$

$$\boldsymbol{w}_Q = \sum_j \beta_j \boldsymbol{q}_j \tag{23}$$

and

$$\boldsymbol{w}_U = \sum_k \gamma_k \boldsymbol{z}_k \ . \tag{24}$$

The following are necessary conditions for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to be optimal for (21)

$$\begin{aligned}
\sum_{i=1}^{m} y_i \alpha_i &= 0 \\
\alpha_i &\le C\,J(i) \quad (i = 1, \ldots, m) \\
\beta_i &\le C\,I(j) \quad (i = 1, \ldots, m) \\
\gamma_k &\le \alpha_i \quad (k = 1, \ldots, \ell,\ i \in I_k) \\
\gamma_k &\le \beta_j \quad (k = 1, \ldots, \ell,\ j \in J_k)
\end{aligned} \tag{25}$$

If the latter hold, then the optimal values of $\boldsymbol{\eta}$, $\boldsymbol{\zeta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ yield the following:

$$\sum_i \eta_i (C\,J(i) - \alpha_i) = \sum_j \zeta_i (C\,I(j) - \beta_i) = \sum_{k=1}^{\ell} \sigma_k \Bigl(\sum_{i \in I_k} \alpha_i - \gamma_k \Bigr) = \sum_{J_k \ne \emptyset} \tau_k \Bigl(\sum_{j \in J_k} \beta_j - \gamma_k \Bigr) = 0\ . \tag{26}$$

It follows that the problem (21) is equivalent to the following dual problem:

$$\begin{aligned}
\text{Minimize} \quad & \tfrac{1}{2} \sum_{i,i'} y_i y_{i'} \boldsymbol{p}_i^\top \boldsymbol{p}_{i'} \alpha_i \alpha_{i'} + \tfrac{1}{2} \sum_{j,j'} \boldsymbol{q}_j^\top \boldsymbol{q}_{j'} \beta_j \beta_{j'} + \tfrac{1}{2} \sum_{k,k'} \boldsymbol{z}_k^\top \boldsymbol{z}_{k'} \gamma_k \gamma_{k'} - \sum_{i=1}^{m} \gamma_i \\
\text{subject to} \quad & \sum_{i=1}^{m} y_i \alpha_i = 0 \\
& 0 \le \alpha_i \le C\,J(i) \quad (i = 1, \ldots, m) \\
& 0 \le \beta_i \le C\,I(j) \quad (j = 1, \ldots, n) \\
& 0 \le \gamma_k \le \alpha_i \quad (k = 1, \ldots, \ell,\ i \in I_k) \\
& 0 \le \gamma_k \le \beta_j \quad (k = 1, \ldots, \ell,\ j \in J_k)
\end{aligned} \tag{27}$$

Note that the size of the latter is *linear*.

3 Extension to nonlinear classification

In the standard formulation of the nonlinear SVM problem, the vectors $\boldsymbol{x}_i$ are lifted to a higher-dimensional space $\Re^M$ by a nonlinear transformation Φ, and the problem is then handled as a linear SVM with examples $\Phi(\boldsymbol{x}_i)$. The dual problem is:

$$\begin{aligned}
\text{Minimize}_{\boldsymbol{\alpha}} \quad & \tfrac{1}{2} \sum_{ij} y_i y_j \Phi(\boldsymbol{x}_i)^\top \Phi(\boldsymbol{x}_j) \alpha_i \alpha_j - \sum_i \alpha_i \\
\text{subject to} \quad & \sum_{i=1}^{m} y_i \alpha_i = 0 \\
& 0 \le \alpha_i \le C\ .
\end{aligned} \tag{28}$$

and the primal solution vector $\boldsymbol{w} \in \Re^M$ must satisfy

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \Phi(\boldsymbol{x}_i) \ . \tag{29}$$

The products $\Phi(\boldsymbol{x}_i)^\top \Phi(\boldsymbol{x}_j)$ can be generated by kernels $K(\boldsymbol{x}, \boldsymbol{x}')$:

$$\psi(\boldsymbol{x}_i)^\top \Phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \ . \tag{30}$$

For example, the so-called quadratic kernel

$$\begin{aligned} K(\boldsymbol{x}, \boldsymbol{x}') &\equiv (\boldsymbol{x}^\top \boldsymbol{x}' + 1)^2 \\ &= (\boldsymbol{x}^\top \boldsymbol{x}')^2 + 2\boldsymbol{x}^\top \boldsymbol{x}' + 1 \\ &= \left(\sum_i x_i x_i' \right)^2 + 2 \sum_i x_i x_i' + 1 \\ &= \sum_i x_i^2 (x_i')^2 + \sum_{i \neq j} x_i x_j x_i' x_j' + 2 \sum_i x_i x_i' + 1 \end{aligned}$$

implements the transformation

$$\Phi(\boldsymbol{x}) = (1, \sqrt{2} x_1, \ldots, \sqrt{2} x_d, x_1^2, \ldots, x_d^2, x_1 x_2, \ldots, x_1 x_d, x_2 x_1, \ldots, x_2 x_3, \ldots, x_2 x_d, \ldots) \tag{31}$$

so that the product $\Phi(\boldsymbol{x}_i)^\top \Phi(\boldsymbol{x}_j)$ can be calculated without calculating the individual values $\Phi(\boldsymbol{x}_i)$ and $\Phi(\boldsymbol{x}_j)$.

3.1 The kernel trick in a join of two tables

In the case of a join of two tables, the examples

$$\boldsymbol{x}_{ij}^\top = (\boldsymbol{p}_i^\top, \boldsymbol{u}_i^\top, \boldsymbol{q}_j^\top)$$

give rise to the following objective function:

$$\tfrac{1}{2} \sum_{i,i'} y_i y_{i'} \boldsymbol{p}_i^\top \boldsymbol{p}_{i'} \alpha_i \alpha_{i'} + \tfrac{1}{2} \sum_{j,j'} \boldsymbol{q}_j^\top \boldsymbol{q}_{j'} \beta_j \beta_{j'} + \tfrac{1}{2} \sum_{k,k'} \boldsymbol{z}_k^\top \boldsymbol{z}_{k'} \gamma_k \gamma_{k'} - \sum_{i=1}^{m} \gamma_i \ . \tag{32}$$

It follows that the linear model can be extended into a (separable) nonlinear one as follows. We consider lifting transformations Φ that preserve the column structure of the table in the sense that for $\boldsymbol{x} = (\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{q})$,

$$\Phi(\boldsymbol{x}) = (\Phi_P(\boldsymbol{p}), \Phi_U(\boldsymbol{u}), \Phi_Q(\boldsymbol{q})) \ .$$

Thus,

$$\Phi(\boldsymbol{x}_{ij})^\top \Phi(\boldsymbol{x}_{i'j'}) = \Phi_P(\boldsymbol{p}_i)^\top \Phi_P(\boldsymbol{p}_{i'}) + \Phi_U(\boldsymbol{u}_i)^\top \Phi_U(\boldsymbol{u}_{i'}) + \Phi_Q(\boldsymbol{q}_i)^\top \Phi_Q(\boldsymbol{q}_{i'}) \ .$$

It follows that our problem 27 can be solved in the higher-dimensional space by modifying the objective function into the following:

$$\tfrac{1}{2} \sum_{i,i'} y_i y_{i'} \Phi_P(\boldsymbol{p}_i)^\top \Phi_P(\boldsymbol{p}_{i'}) \alpha_i \alpha_{i'} + \tfrac{1}{2} \sum_{j,j'} \Phi_Q(\boldsymbol{q}_j)^\top \Phi_Q(\boldsymbol{q}_{j'}) \beta_j \beta_{j'} + \tfrac{1}{2} \sum_{k,k'} \Phi_U(\boldsymbol{z}_k)^\top \Phi_U(\boldsymbol{z}_{k'}) \gamma_k \gamma_{k'} - \sum_{i=1}^{m} \gamma_i \ . \tag{33}$$

and the "kernel trick" can be applied if we use transformations that are consistent with conventional kernels, $K_P(\boldsymbol{p}, \boldsymbol{p}') = \Phi_P(\boldsymbol{p})^\top \Phi_P(\boldsymbol{p}')$, $K_U(\boldsymbol{u}, \boldsymbol{u}') = \Phi_U(\boldsymbol{u})^\top \Phi_U(\boldsymbol{u}')$ and $K_Q(\boldsymbol{q}, \boldsymbol{q}') = \Phi_Q(\boldsymbol{q})^\top \Phi_Q(\boldsymbol{q}')$, so the objective can be evaluated in the original space.