Abstract
In the present paper we describe a new algorithm for Support Vector regression (SVR). Like standard $\nu$-SVR algorithms, this algorithm automatically adjusts the radius of insensitivity (tube width $\varepsilon$) to fit the data. However, this is achieved without additional complexity in the optimisation problem. Moreover, by careful modification of the kernel function, we are able to significantly simplify the form of the dual SVR optimisation problem.

INTRODUCTION
Support Vector Machines are a relatively new class of learning machines that have evolved from the concepts of structural risk minimisation (SRM) (in the pattern recognition case) and regularisation theory (in the regression case). The major difference between Support Vector Machines (SVMs) and many other Neural Network (NN) approaches is that instead of tackling problems using the traditional empirical risk minimisation (ERM) method, SVMs use the concept of regularised ERM. This has enabled people to use SVMs with potentially huge capacities on smaller datasets without running into the usual difficulties of overfitting and poor generalisation performance.

Geometrically, the basis of SVM theory is to nonlinearly map the input data into some (possibly infinite dimensional) feature space where the problem may be treated as a linear one. In particular, when tackling regression problems using SVMs, the output is a linear function of position in feature space. However, the complexities of this feature space (and the non-linear map associated with it) are “hidden” using a kernel function. It is this ability to hide complexity (resulting in a simple linearly constrained N-dimensional quadratic programming problem with no non-global minima), along with the ability to use complex models while avoiding overfitting, that has made SVM methods so popular over recent years.

In the present paper, we give a new formulation for the SVM regression problem that, while incorporating many of the features of existing methods (in particular, the $\nu$-Support Vector Regression ($\nu$-SVR) methods), dispenses with much of the complexity that comes with these formulations. Specifically, by making the primal cost function entirely quadratic in nature, we are able to reduce the SVM training problem to a particularly simple quadratic programming problem with a unique global minima whose only constraints are the positivity of all variables.

$\varepsilon$-SVR REGRESSION
The standard SV regression problem is formulated as follows. Suppose we are given a training set:

$$\Theta = \{(x_1, z_1), (x_2, z_2), \ldots, (x_N, z_N)\}$$

where:

$$x_i \in \mathbb{R}^{d_L},$$

$$z_i \in \mathbb{R}$$

which is assumed to have been generated based on some unknown but well defined map:

$$\hat{y}_i = \hat{g}(\hat{x}_i) + \text{noise}$$

We (implicitly, as will be seen later) define a set of functions $\varphi_j : \mathbb{R}^{d_L} \rightarrow \mathbb{R}$, $1 \leq j \leq d_H$, which collectively make up a map from input space to feature space, namely $\varphi : \mathbb{R}^{d_L} \rightarrow \mathbb{R}^{d_H}$, where:

$$\varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \\ \varphi_{d_H}(x) \end{bmatrix}$$

Using the functions $\varphi_j : \mathbb{R}^{d_L} \rightarrow \mathbb{R}$, $1 \leq j \leq d_H$, we aim to find a non-linear approximation to $\hat{g}$:

$$g(x) = w^T \varphi(x) + b$$

which is a linear function of position in feature space. The usual $\varepsilon$-SVR method of selecting $w$ and $b$ is to minimise the regularised risk functional:

$$\min_{w, b, \xi, \xi^*} \quad R(w, b, \xi, \xi^*) = \frac{1}{2}w^T w + \frac{C_1}{2} \xi^T \xi + \frac{C_2}{2} 1^T \xi + \frac{1}{N} 1^T \xi^*$$

such that:

$$- \frac{1}{2} (w^T \varphi(x_i) + b) \geq z_i - \varepsilon - \xi_i$$

$$- (w^T \varphi(x_i) + b) \geq - z_i - \varepsilon - \xi^*_i$$

$$\xi, \xi^* \geq 0$$

where $\frac{1}{2} w^T w$ characterises the complexity of the model (the larger $\frac{1}{2} w^T w$, the larger the gradient of $g(x)$ in feature space, and hence the more $g(x)$ may vary for a given variation in input, $x_i$); and $\frac{1}{2} 1^T \xi + \frac{1}{N} 1^T \xi^*$ the empirical risk associated with it. The constant $C > 0$ controls the trade-off between empirical risk minimisation (and potential overfitting) if $C$ is large and complexity minimisation (and potential underfitting) if $C$ is small.

The constant $\varepsilon > 0$ term in (1) is included to give the model a degree of noise insensitivity. Essentially, so long as $z_i - \varepsilon \leq g(x_i) \leq z_i + \varepsilon$, $\xi^*_i = \xi_i = 0$, and so there will be no empirical risk associated with small perturbations.
Boundary vectors:

Error vectors: $\alpha$

Noting that each training vector corresponds to one pair of boundary vectors, the process.

One practical difficulty with the $\nu$-SV, which is the complexity of the constraint set, and in particular the presence of the upper bound constraint $1^T (\alpha + \alpha^*) \leq C\nu$. We would like to remove this constraint without losing the ability to automatically select $\epsilon$ based on another, more useful parameter, $\nu$. Consider the primal form of the standard $\nu$-SV regression, (4). The term $C\nu\epsilon$ is effectively a linear regularisation term for the variable $\epsilon$ (in much the same way that $\frac{1}{2} w^Tw$ is a regularisation term for the variable $w$). We replace this linear regularisation term with a quadratic regularisation term, $C\nu\epsilon^2$, to get the new regularised risk functional:

$$
\min_{w,b,\xi,\xi^*,\epsilon} R(w, b, \xi, \xi^*, \epsilon) = \frac{1}{2} w^Tw + C\nu\epsilon^2
$$

such that:

$$
\begin{align*}
\epsilon^T (\xi - \xi^*) & \geq \epsilon_i - \xi_i \\
\epsilon^T (\xi - \xi^*) & \geq -z_i - \epsilon - \xi_i^* \\
\xi, \xi^* & \geq 0
\end{align*}
$$

where, once again, $\nu > 0$ is a constant. Note that this problem does not contain an inequality $\epsilon \geq 0$. However, the optimal solution to (6) will satisfy this constraint. To see

**MODIFIED $\nu$-SV REGRESSION**

One practical difficulty with (5) is the complexity of the constraint set, and in particular the presence of the upper bound constraint $1^T (\alpha + \alpha^*) \leq C\nu$. We would like to remove this constraint without losing the ability to automatically select $\epsilon$ based on another, more useful parameter, $\nu$. Consider the primal form of the standard $\nu$-SV regression, (4). The term $C\nu\epsilon$ is effectively a linear regularisation term for the variable $\epsilon$ (in much the same way that $\frac{1}{2} w^Tw$ is a regularisation term for the variable $w$). We replace this linear regularisation term with a quadratic regularisation term, $C\nu\epsilon^2$, to get the new regularised risk functional:

$$
\min_{w,b,\xi,\xi^*,\epsilon} R(w, b, \xi, \xi^*, \epsilon) = \frac{1}{2} w^Tw + C\nu\epsilon^2
$$

such that:

$$
\begin{align*}
\epsilon^T (\xi - \xi^*) & \geq \epsilon_i - \xi_i \\
\epsilon^T (\xi - \xi^*) & \geq -z_i - \epsilon - \xi_i^* \\
\xi, \xi^* & \geq 0
\end{align*}
$$

where, once again, $\nu > 0$ is a constant. Note that this problem does not contain an inequality $\epsilon \geq 0$. However, the optimal solution to (6) will satisfy this constraint. To see
why this must be the case, suppose that the optimal solution does not satisfy \( \varepsilon \geq 0 \). Given this solution, if we replace \( \varepsilon < 0 \) with \( \varepsilon = 0 \) then clearly the inequality constraints in (6) will still be satisfied. Furthermore, \( R (w, b, \xi, \xi^*, \varepsilon) \leq R (w, b, \xi, \xi^*, 0) \forall \varepsilon \neq 0 \). Therefore any solution not satisfying \( \varepsilon \geq 0 \) is not optimal, so a constraint of the form \( \varepsilon \geq 0 \) would be superfluous.

Following the same approach as used for automatically biased SVMs [2], we may define an augmented feature space

\[
\varphi_A (x, d) = \left[ \varphi (x) \right]
\]

We also define the augmented weight vector:

\[
w_A = \left[ w \varepsilon \right]
\]

Using these definitions, (6) may be re-written thus:

\[
\min_{w, b, \xi, \xi^*} R (w_A, b, \xi, \xi^*) = \frac{1}{2} w_A^T Q w_A \\
\text{such that:} \quad \left( \begin{array}{c}
\sum_i (\alpha_i - \alpha_i^*) \varphi (x_i) \\
\varepsilon = \frac{1}{C\nu} (1^T \alpha + 1^T \alpha^*)
\end{array} \right) \geq z_i - \xi_i \\
\xi, \xi^* \geq 0
\]

where:

\[
Q = \left[ \begin{array}{cc}
1 & 0 \\
0^T & C\nu
\end{array} \right]
\]

To form the dual problem, we introduce the non-negative Lagrange multipliers \( \alpha_i \) and \( \alpha_i^* \) associated with the first two inequality constraints of (7), respectively. Following the usual methods, it is not difficult to show that:

\[
w = \sum_i (\alpha_i - \alpha_i^*) \varphi (x_i) \\
\varepsilon = \frac{1}{C\nu} (1^T \alpha + 1^T \alpha^*)
\]

which leads us to the dual formulation:

\[
\min_{\alpha, \alpha^*} L (\alpha, \alpha^*) = \left[ \begin{array}{c}
\alpha \\
\alpha^*
\end{array} \right]^T \left[ \begin{array}{c}
H \alpha \\
H \alpha^*
\end{array} \right] - \left[ \begin{array}{c}
\alpha \\
\alpha^*
\end{array} \right]^T s \\
\text{such that:} \quad 0 \leq \alpha \leq \frac{C}{\nu} 1 \\
0 \leq \alpha^* \leq \frac{C}{\nu} 1 \\
1^T (\alpha - \alpha^*) = 0
\]

where:

\[
s = \left[ \begin{array}{c}
z \\
-\sum_i \varphi (x_i)
\end{array} \right] \\
H = \left[ \begin{array}{cc}
G_+ & -G_+ \\
-G_+ & G_+
\end{array} \right] \\
G_+ = G + \frac{1}{C\nu} 11^T \\
G_+ = G - \frac{1}{C\nu} 11^T \\
G_{i,j} = K (x_i, x_j)
\]

Considering (9), it should be noted that:

- The hessian matrix \( H \) is positive semi-definite and the constraints are linear. Hence there will be no non-global minima.
- While the modified \( \nu \)-SV regression method incorporates tube-shrinking into its design, the constraint set of (9) is no more complex than the standard \( \varepsilon \)-SVR dual, (2).

We now investigate the effect of the constant \( \nu \). Consider the primal form of the modified \( \nu \)-SVR problem, namely (6).

Suppose we have a solution (and in particular \( \varepsilon, \xi \) and \( \xi^* \)) which we claim is optimal. If this claim is true then:

\[
\xi_i = \begin{cases} 
\frac{g (x_i) - (z_i - \varepsilon)}{\alpha_i} & \text{if } \alpha_i \neq 0 \\
0 & \text{otherwise}
\end{cases} \\
\xi_i^* = \begin{cases} 
\frac{g (x_i) - (-z_i - \varepsilon)}{\alpha_i^*} & \text{if } \alpha_i^* \neq 0 \\
0 & \text{otherwise}
\end{cases}
\]

Furthermore, any change in \( \varepsilon \) (and requisite modification of \( \xi \) and \( \xi^* \) to satisfy (10)) should lead to an increase in \( R (w, b, \xi, \xi^*, \varepsilon) \). Suppose we increase \( \varepsilon \) by some very small amount, \( \Delta \varepsilon > 0 \), while keeping \( w \) and \( b \) constant, and changing \( \xi \) and \( \xi^* \) appropriately to satisfy (10). The change in \( R (w, b, \xi, \xi^*, \varepsilon) \) will be, to first order:

\[
\Delta R = \left( \frac{\partial R}{\partial \varepsilon} = \frac{\partial R}{\partial \varepsilon} \right) \Delta \varepsilon
\]

To summarise, for the modified \( \nu \)-SV regressor described in this section, the following bounds hold:

- \( \varepsilon \geq \frac{1}{\nu} \frac{N \nu}{\nu} \).
- \( \varepsilon \leq \frac{1}{\nu} \frac{N \nu}{\nu} \).

The constant \( \nu \) controls a bounding relation between the tube-width \( \varepsilon \) and the fraction of support and error vectors.

**FURTHER SIMPLIFICATIONS**

**Automatic Biasing**

We can remove the equality constraint in (9) by using the automatic biasing method of [3], [2]. We extend (6) by adding an additional term, \( \frac{1}{2} b^2 \), to the regularised risk functional to get:

\[
\min_{w, b, \xi, \xi^*, \varepsilon} R (w, b, \xi, \xi^*, \varepsilon) = \frac{1}{2} w^T w + \frac{C}{\nu} \varepsilon^2 + \frac{1}{2} b^2 \\
\text{such that:} \quad \left( \begin{array}{c}
w^T \varphi (x_i) + b \\
\varphi (x_i) \end{array} \right) \geq z_i - \varepsilon - \xi_i \\
\xi, \xi^* \geq 0
\]

(11)
where $\beta > 0$ is a constant. Our augmented feature map becomes:

$$
\varphi_A (x, d) = \begin{bmatrix} \varphi (x) \\ d \\ 1 \end{bmatrix}
$$

and likewise the augmented weight vector is:

$$
w_A = \begin{bmatrix} w \\ \varepsilon \\ b \end{bmatrix}
$$

Hence we can re-write (11) thusly:

$$
\min_{w_A, b, \xi, \xi^*} R (w_A, b, \xi, \xi^*) = \frac{1}{2} w_A^T Q w_A + \frac{C}{N} (1^T \xi + 1^T \xi^*)
$$

such that:

$$
-w_A^T \varphi_A (x_i, +1) \geq z_i - \xi_i
$$

$$
-w_A^T \varphi_A (x_i, -1) \geq -z_i - \xi_i^*
$$

where:

$$
\xi, \xi^* \geq 0
$$

Using these definitions, (6) may be re-written as:

$$
\min_{w_A} R (w_A) = \frac{1}{2} w_A^T Q w_A
$$

such that:

$$
-w_A^T \varphi_A (x_i, +1, i) \geq z_i
$$

$$
-w_A^T \varphi_A (x_i, -1, 0, i) \geq -z_i
$$

where:

$$
w_A = \begin{bmatrix} w \\ \varepsilon \\ b \end{bmatrix}
$$

and hence $e_0 = 0$. We also define the augmented weight vector:

$$
w_A = \begin{bmatrix} w \\ \varepsilon \\ b \end{bmatrix}
$$

Note that (14) contains no lower bound on $\xi, \xi^*$. This is unnecessary for the same reason that a lower bound on $\varepsilon$ is unnecessary (i.e. such a constraint would be superfluous). In this case, our augmented feature map is:

$$
\varphi_A (x, d, i, j) = \begin{bmatrix} \varphi (x) \\ d \\ e_i \\ e_j \end{bmatrix}
$$

where the vector $e_i \in \mathbb{R}^N$ consists of all zero elements except for the $i^{th}$ element, which is 1 (and hence $e_0 = 0$). We also define the augmented weight vector:

$$
w_a = \begin{bmatrix} w \\ \varepsilon \\ b \end{bmatrix}
$$

Following the usual method, we find that:

$$
w = \sum_i (\alpha_i - \alpha_i^*) \varphi (x_i)
$$

$$
\varepsilon = \frac{1}{\sqrt{\nu}} \left( 1^T \alpha + 1^T \alpha^* \right)
$$

$$
b = \frac{1}{2} \left( 1^T \alpha - 1^T \alpha^* \right)
$$

Hence the dual problem becomes:

$$
\min_{\alpha, \alpha^*} L (\alpha, \alpha^*) = \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right]^T H \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right] - \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right]^T s
$$

such that:

$$
0 \leq \alpha \leq \frac{C}{N} 1
$$

$$
0 \leq \alpha^* \leq \frac{C}{N} 1
$$

where $s$ is unchanged from the previous section and:

$$
H = \begin{bmatrix} G_+ & -G_- \\ -G_+ & G_- \end{bmatrix}
$$

$$
G_+ = G + \left( \frac{1}{\nu} + \frac{1}{\beta} \right) 11^T
$$

$$
G_- = G - \left( \frac{1}{\nu} - \frac{1}{\beta} \right) 11^T
$$

$$
G_{i,j} = K (x_i, x_j)
$$

which is essentially the same as (9), except that there is no equality constraint present.

**Quadratic Empirical Risk**

Finally, we may remove the upper bound constraints on $\alpha$ and $\alpha^*$ while at the same time making $H$ positive definite by modifying the form of the empirical risk component of our regularised risk functional as follows:

$$
\min_{w, b, \xi, \xi^*, \varepsilon} R (w, b, \xi, \xi^*, \varepsilon) = \frac{1}{2} w^T w + \frac{C}{N} \varepsilon^2 + \frac{C}{N} \xi^T G_+ \xi + \frac{C}{N} \xi^T G_- \xi^*
$$

such that:

$$
w^T \varphi (x_i) + b \geq z_i - \varepsilon - \xi_i
$$

$$
w^T \varphi (x_i) + b \geq -z_i - \varepsilon - \xi_i^*
$$

Note that (14) contains no lower bound on $\xi, \xi^*$. This is unnecessary for the same reason that a lower bound on $\varepsilon$ is unnecessary (i.e. such a constraint would be superfluous). In this case, our augmented feature map is:

$$
\varphi_A (x, d) = \begin{bmatrix} \varphi (x) \\ d \\ 1 \end{bmatrix}
$$

$$
\varphi_A (x, d, i, j) = \begin{bmatrix} \varphi (x) \\ d \\ e_i \\ e_j \end{bmatrix}
$$

where the vector $e_i \in \mathbb{R}^N$ consists of all zero elements except for the $i^{th}$ element, which is 1 (and hence $e_0 = 0$). We also define the augmented weight vector:

$$
w_A = \begin{bmatrix} w \\ \varepsilon \\ b \end{bmatrix}
$$

Following the usual method, we find that:

$$
w = \sum_i (\alpha_i - \alpha_i^*) \varphi (x_i)
$$

$$
\varepsilon = \frac{1}{\sqrt{\nu}} \left( 1^T \alpha + 1^T \alpha^* \right)
$$

$$
b = \frac{1}{2} \left( 1^T \alpha - 1^T \alpha^* \right)
$$

$$
\xi = \frac{C}{\sqrt{\nu}} \alpha
$$

$$
\xi^* = \frac{C}{\sqrt{\nu}} \alpha^*
$$

The dual problem becomes:

$$
\min_{\alpha, \alpha^*} L (\alpha, \alpha^*) = \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right]^T H \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right] - \left[ \begin{array}{c} \alpha \\ \alpha^* \end{array} \right]^T s
$$

such that:

$$
\alpha \geq 0
$$

$$
\alpha^* \geq 0
$$

where $s$ is unchanged from the previous section and:

$$
H = \begin{bmatrix} G_+ & -G_- \\ -G_+ & G_- \end{bmatrix}
$$

$$
G_+ = G + \left( \frac{1}{\nu} + \frac{1}{\beta} \right) 11^T + \frac{N}{\sqrt{\nu}} I
$$

$$
G_- = G - \left( \frac{1}{\nu} - \frac{1}{\beta} \right) 11^T
$$

$$
G_{i,j} = K (x_i, x_j)
$$

which is essentially the same as (13), except that there are no upper bounds present and, furthermore, $H$ is positive.
definite, implying that there is a unique global solution. One possible disadvantage of this form is a slight decrease in the sparsity of the solution if the training set is degenerate. The bounds $\varepsilon > \frac{1}{N \nu} \sigma$ and $\varepsilon < \frac{1}{N \nu} \sigma$ no longer apply to this formulation. Both are superseded by a stricter, equality relation. Before proceeding to derive this relation, we must first define the following quantity:

$$\bar{e} = \frac{1}{N} \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

$\bar{e}$ is the mean error (distance outside the $\varepsilon$ tube) of all training points (including non-error points, which are defined as lying at a distance of 0 from the $\varepsilon$-tube). Using (17), it is easy to see that the following relation will hold for the form of modified $\nu$-SV regressor described in the present section:

$$\varepsilon = \frac{1}{\nu} \bar{e}$$

So, for modified $\nu$-SV regressors using quadratic empirical risk, as described in the present section, there is a direct proportional relationship between the mean error of all training points and the width of the $\varepsilon$-tube. Furthermore, the constant of proportionality may be set exactly using the training constant $\nu$.

**EXPERIMENTAL RESULTS**

All code for this experiment was written in C++ and compiled using DJGPP on a 1GHz Pentium III with 512MB of memory, running Windows 2000. When considering modified $\nu$-SVRs we have exclusively used form (16) (auto-biasing, quadratic empirical risk) for reasons of implementational simplicity.

To evaluate the effectiveness of our algorithm, following [5] we have chosen the task of using regression to estimate a noisy sinc function from $N$ samples $(x_i, z_i)$, where $x_i$ is drawn uniformly from $[-3, 3]$, $z_i = \frac{\sin(\pi x_i)}{\pi x_i} + v_i$, and $v_i$ is drawn from a Gaussian with zero mean and variance $\sigma^2$.

Unless otherwise stated, $N = 50, C = 100, \sigma = 0.2$ and $K(x,y) = e^{-||x-y||^2}$.

Figures 1 and 2 show the results achieved using a modified $\nu$-SVR with $\nu = 0.2$ and $\nu = 0.8$, respectively. When $\nu = 0.2$, the algorithm automatically sets $\varepsilon = 0.17$, and when $\nu = 0.8$, we find $\varepsilon = 0.09$. This demonstrates the similarities between the role of $\nu$ in standard and modified $\nu$-SVR methods.

Figures 3 and 4 show the results achieved using a modified $\nu$-SVR with $\nu = 0.2$ when dealing with training data with noise $\sigma = 0$ (figure 3) and $\sigma = 1$ (figure 4). In both cases, the tube width $\varepsilon$ is able to adjust automatically to fit the noise present in the problem, just as occurs in the standard $\nu$-SVR approach. For comparison, figures 5 and 6 show the result achieved using a standard $\varepsilon$-SVR method with $\varepsilon = 0.2$, which is not optimal in either case. In figure 5, the estimate is biased, and in figure 6, $\varepsilon$ does not match the noise (c.f. [5]).

**Discussion**
From these results, and [5], it would appear that $\nu$-SVR and modified $\nu$-SVR methods exhibit many similar characteristics, and produce results that are essentially identical to one another. It is reasonable then to ask why we feel it is reasonable to suggest modifying $\nu$-SVR methods in the first place?

The answer, we believe, lies in the relative simplicity of the dual forms of the modified $\nu$-SVR methods (9), (13) and (16) when compared with the standard $\nu$-SVR method (5). It is our experience that the presence of the inequality constraint $\mathbf{1}^T(\alpha + \alpha^*) \leq C\nu$ significantly complicates the task of solving the dual optimisation problem. This difficulty is particularly noticeable when attempting to develop an incremental approach to the problem of $\nu$-SV regression.

Our results suggest that it is possible to achieve results comparable with those achieved using $\nu$-SVR methods without the added computational complexity typically associated with such methods.

**CONCLUSION**

We have presented a modified $\nu$-SVR algorithm which, like standard $\nu$-SVR methods, is able to automatically select the tube with $\varepsilon$. However, we have achieved this without the additional computational complexity usually associated with $\nu$-SVR methods. We have demonstrated on a sample problem that our method is able to achieve results which are essentially identical to the standard $\nu$-SVR method, demonstrating that this loss of computational complexity has not been accompanied by any significant degradation in regressor performance.

**REFERENCES**

Author/s:
Shilton, A.; Palaniswami, M.

Title:
Modified nu-SV Method for Simplified Regression

Date:
2004

Citation:

Publication Status:
Published

Persistent Link:
http://hdl.handle.net/11343/34062

File Description:
Modified nu-SV Method for Simplified Regression

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.