Analyzing the usefulness of SVMs v. RVMs when predicting breast cancer

Harsh Vardhan Pachisia
University of Chicago
harshpachisia@uchicago.edu

Shwetha Srinivasan
University of Chicago
shwetha@uchicago.edu

Jonathan Juarez
University of Chicago
jonathanjuarez@uchicago.edu

Abstract

This project aimed to investigate and compare the performance of Support Vector Machines (SVMs) and Relevance Vector Machines (RVMs) in the classification of benign and malignant breast tumors. We introduce the mechanisms of how SVMS's work, their usefulness, and potential drawbacks and analyze why RVM in classifying tumors can lead to better results. In the end, comparing several models determine that an RVM model gives greater accuracy scores compared to SVM model.

1 Introduction

This project aims to investigate and compare the performance of Support Vector Machines (SVMs) and Relevance Vector Machines (RVMs) in the classification of benign and malignant breast tumors. The dataset under consideration is sourced from UCI Machine Learning, comprising a featurized set with 10 real-value features computed for each cell nucleus. This course introduces the SVM model, and while useful has some drawbacks that RVM models aim to solve.

Our methodology includes conducting an in-depth literature review of SVMs and RVMs, focusing on their underlying principles, strengths, and weaknesses. Next, we analyze the breast cancer dataset to comprehend its structure, features and target variables. Third, we implement the SVM and RVM based on lessons learnt in class and outside literature review. Finally, we evaluate their performance using relevant metrics and conduct a comparative analysis looking at interpretability, robustness, etc. Ultimately, we aim to shed light on how RVMs compare with SVMs in breast cancer classification.

2 Literature Review

During this course, we have been introduced to the mechanisms of how SVM's work, their usefulness, and potential drawbacks. In particular, the SVM suffers from the fact that it does not have probabilistic outputs, the requirement to estimate a trade-off parameter, and the need to utilise 'Mercer' kernel functions. The RVM model was introduced by Tipping, 1999 under Microsoft Research to mitigate some of these drawbacks. We want to build upon and evaluate whether RVMs can mitigate some of the drawbacks of SVMs using the breast cancer classification dataset. RVMs have been used for both regression and classification tasks such as target detection in images (regression) and classification of microcalcifications from mammograms (classification) (Tzikas et.al, 2006).

Tipping uses synthetic examples such as the sin function and Gaussian mixture data to demonstrate the reduction in complexity for a RVM model in comparison to SVM. For a real world dataset, such

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

as the Pima Indian diabetes dataset, the number of kernels reduced from 109 to 4 whereas they declined from 2540 to 316 kernels for U.S.P.S handwritten digit dataset (Tipping, 2001). While SVM performed better for the U.S.P.S dataset and RVM had better accuracy for the Pima dataset, there wasn't a significant difference between the two in terms of accuracy. Similar to these examples, we hope to demonstrate that applying these two models in the breast cancer dataset for classification of benign and malignant tumors, we would notice a reduction in complexity with the RVM model without detracting from its performance in classification tasks.

3 Data sources and Features Explanation

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

Attribute Information:

1) ID number 2) Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter) b) texture (standard deviation of gray-scale values) c) perimeter d) area e) smoothness (local variation in radius lengths) f) $\frac{\text{perimeter}^2}{\text{area}-1.0}$

g) concavity (severity of concave portions of the contour) h) concave points (number of concave portions of the contour) i) symmetry j) fractal dimension ("coastline approximation" - 1)

The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

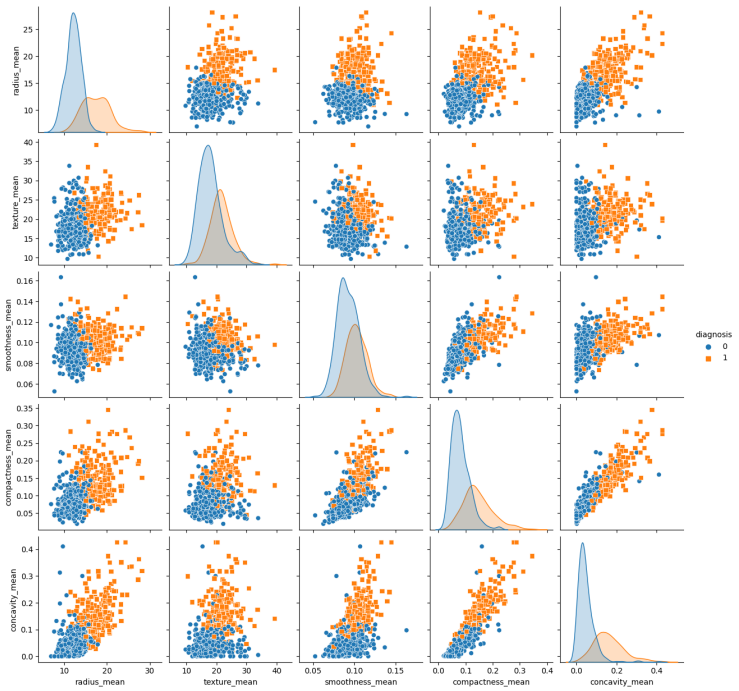


Figure 1: Correlation of Feature Variables

4 The SVM and RMV models

We define the following functions that will be compared below. These are the functions used in the Python scripts that generated our models.

4.1 SVM with RBF Kernel

SVM Model with RBF (Radial Basis Function) Kernel:

$$\begin{aligned} \text{Given training data:} \quad & \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \\ \text{where } x_i \quad & \text{is a feature vector, and } y_i \in \{0, 1\} \text{ is the class label.} \\ \text{The decision function is:} \quad & f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \\ \text{where } K(x_i, x) \quad & \text{is the kernel function (RBF in this case),} \\ & \text{and } \alpha_i \text{ are the regularization parameters.} \\ \text{This is defined as:} \quad & K(x_i, x_j) = e^{-\frac{1}{n_{\text{features}}} \|x_i - x_j\|^2} \end{aligned} \tag{1}$$

The exponential RBF $e^{-\frac{1}{n_{\text{features}}} \|x_i - x_j\|^2}$ transforms the distance into a similarity measure. A smaller distance results in a higher similarity. $\frac{1}{n_{\text{features}}}$ is the scaling factor, where n_{features} is the number of features in the dataset. $\|x_i - x_j\|^2$ represents the squared Euclidean distance between the feature vectors x_i and x_j.

The use of RBF kernels in vector machines is a common choice due to several reasons, including their ability to capture complex, non-linear relationships in the data [2]. RBF kernels are especially suitable when the decision boundary is not linear and exhibits intricate patterns. From the breast cancer data in using 10 various features and visually inspecting the plots of of tumor classification, this is a case that calls for the usage of a non-linear kernel

The key advantage of RBF kernels lies in their flexibility to map the input space into a higher-dimensional space, making it easier to separate classes or model complex relationships, which is known as the "kernel-trick".

4.2 RVM with RBF Kernel

The Relevance Vector Machine (RVM) model, particularly for classification, can be summarized with the following equations and explanations. These models are similar to the models directly used by James Ritchie, who directly used formulations written by Bishop and Nasrabadi [3], broken down into several steps.

Decision Function:

$$f(x) = \sum_{i=1}^{n} \beta_i K(x_i, x) \tag{2}$$

$f(x)$ is a weighted sum of kernel evaluations. β_i is the weight associated with the i-th data point. $K(x_i, x)$ is the kernel function (again the radial basis function, RBF), computing the similarity between data point x and the i-th support vector x_i.

Logistic Function (Sigmoid):

$$P(y = 1|x) = \frac{1}{1 + e^{-f(x)}} \tag{3}$$

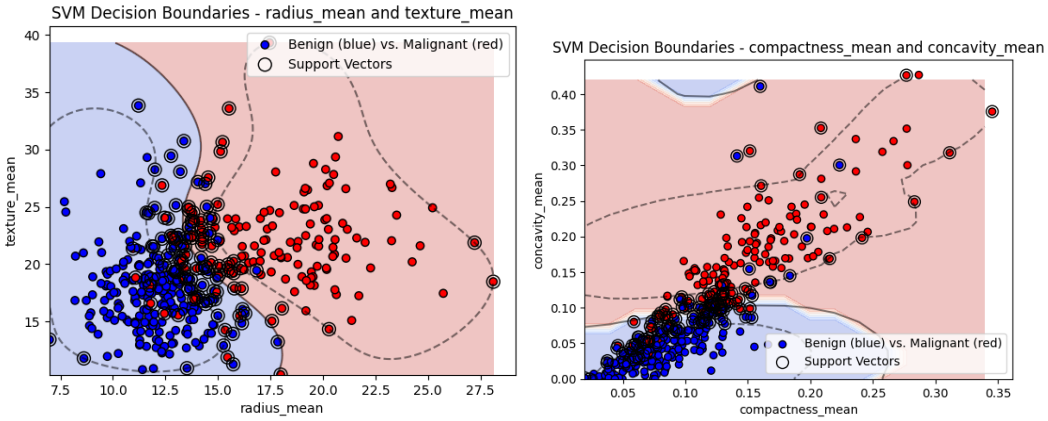
$P(y = 1|x)$ is the probability that the output y is 1 given input x.

Negative Log-Likelihood:

$$\text{Negative Log-Likelihood} = -\left(\sum_{i=1}^{N} \log(P(y_i|x_i)) + \frac{1}{2} \sum_{i=1}^{n} \log(\alpha_i) \right) \tag{4}$$

This represents the negative log-likelihood of the data. $\sum_{i=1}^{N} \log(P(y_i|x_i))$ is the log-likelihood of the observed data given the model using Baye's rule. $\frac{1}{2} \sum_{i=1}^{n} \log(\alpha_i)$ penalizes model complexity through the regularization term. α_i is a regularization parameter associated with the i-th data point. This function is being optimized in our RVM model to classify tumors.

5 Results



(a) SVM using radius and texture

(b) SVM using compactness and concavity

Figure 2: SVM using two features

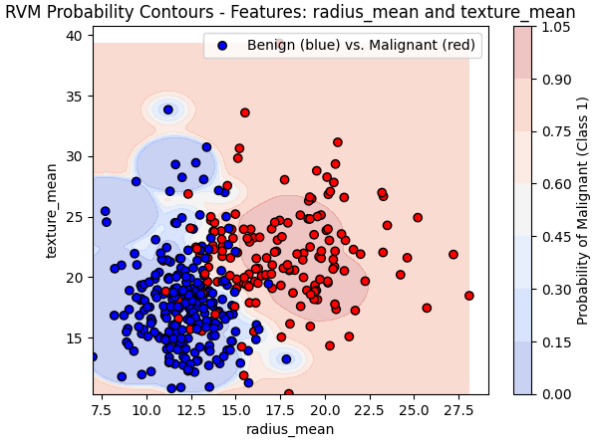


Figure 3: RVM Probability Contours

The probability contours from the RVM model represent regions of the input space where the model assigns different probabilities to different classes. Each point in the input space is associated with a probability of belonging to a particular class.

In the case of benign vs. malignant tumors, the contours represent the decision boundaries between the two classes. The contours indicate the regions where the RVM model assigns a higher probability to one class compared to the other. The specific values on the contours represent the probability of belonging to the positive class. This demonstrates a more nuanced view of the model's predictions compared to a strict decision boundary.

Model performance: The Jaccard Score is a measure of similarity between two sets. In the context of classification, it is often used to measure the similarity between the predicted set of labels and the true set of labels. The Jaccard Score is defined as the size of the intersection divided by the size of the union of the sample sets. The Jaccard score for the SVM model is approximately 76% whereas the score for the RVM model is approximately 78%.

6 Conclusion

While the difference in the Jaccard score is minimal between the two models, the RVM model does offer some advantages in terms of utilizing fewer kernel functions. Sparsity achieved through RVM has important considerations in practical applications as it saves memory and computation. In the health field, particularly when predicting cancer diagnoses for patients, it is crucial to carefully evaluate and choose the most effective models to ensure the best possible outcomes.

References

[1] Tipping, Michael. "The relevance vector machine." Advances in neural information processing systems 12 (1999).

[2]Burges, Christopher JC. "A tutorial on support vector machines for pattern recognition." Data mining and knowledge discovery 2.2 (1998): 121-167.

[3] Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4. No. 4. New York: springer, 2006.

[4]Wolberg,William, Mangasarian,Olvi, Street,Nick, and Street,W. (1995). Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B.

[5] Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. Journal of machine learning research, 1(Jun), 211-244.

[6]Tzikas, Dimitris & Wei, Liyang & Likas, Aristidis & Yang, Yongyi. (2006). A tutorial on relevance vector machines for regression and classification with applications. Eurosip. 17.