جامعة الموصل
كلية علوم الحاسوب والرياضيات

توظيف خوارزمية المفترسات البحرية لتحسين آلةُ المُتجهِ الدَاعِمِ الجَّزائيَّةِ Lp في التنصيفِ الثنائيِّ ذي الأَبعادِ العَاليَةِ

فهد علي إبراهيم عزب

رسالة ماجستير

الإحصاء

بإشراف

المدرس

د. نعم عبد المنعم عبد المجيد الذنون

1443هـ 2021م

المستخلص

أصبحت البيانات ذات الأَبعاد العالية دوراً متزايداً في المعلوماتية الحياتية التي يفوق عدد سماتها (المتغيرات) عدد المشاهدات ممَّا يجعل العديد من الطرائق والأَساليب الإحصائية صعبة التطبيق . اكتسبت الطرائق الجزائية أهمية كبيرة بين الإحصائيين بوصفها مفتاحًا لأداء اختيار الصفات (المتغيرات) وتقدير الأنموذج في وقت واحد. يعد موضوع آلة المتجه الداعم من المواضيع المستخدمة بكثرة في العديد من المجالات التطبيقية بالرغم من عدم قدرته على أداء اختيار الصفات. يعدُّ موضوع الطرائق الجزائية ذا أهمية لاختيار المتغيرات في التصنيف الثنائي ذي الأَبعاد العالية والتي يكون حجم المتغيرات كبير جدًا مما يؤثر على جودة التصنيف الثنائي. يصنف أُسلوب آلة المتجه الداعم الجزائية واحدًا من أَفضل الأَساليب المستخدمة بسبب أَدائه في اختيار الصفات ولأجرائه التصنيف في آن واحد. يُعدُّ حد الجزاء Lp-norm من أَكثر حدود الجزاء استخدامًا، اذ يعتمد اسلوب آلة المتجه الداعم الجزائية على قيمة معلمة الضبط p ، فأختيار يؤدِّي دورًا مميزًا في جودة أُسلوب آلة المتجه الداعم الجزائية. تم في هذه الرسالة استخدام دالة الجزاء Lp مع آلة المتجه الداعم ، واقترح توظيف خوارزمية المفترسات البحرية لغرض تقليل حجم المتغيرات بهدف الحصول على تصنيف عالي ، وللتحقق من كفاءة الخوارزمية المقترحة تم اختبار كفاءتها على بيانات ومقارنتها مع أُسلوب العبور المتقاطع Cross Validation والتي تكتب أختصاراً (CV) من خلال إجراء تجارب محاكاة والتطبيق على بيانات حقيقية تتناول عدد من المجالات العلمية. أَظهرت النتائج تفوق الخوارزمية المقترحة على أُسلوب CV من خلال اختيار افضل السمات وبأعلى تصنيف .

UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS

Employing the marine predators algorithm to improve the Penalized support vector machine Lp penalty in high-dimensional binary classification

Fahad Ali Abrahim Al Azab

M.Sc./Thesis

Statistics

Supervised by

Lecturer
Dr. Niam Abdulmunim Abdulmajeed Al-Thanoon

2021 A.D. **1443 A.H.**

ABSTRACT

High-dimensional data has become increasingly present in bioinformatics , in which the number of features (variables) exceeds the number of observations, making many statistical methods and methods difficult to apply and use. In recent years, penalized methods have gained great importance among statisticians as a key to performing simultaneous feature selection (variables) and model estimation (classification). The subject of the support vector machine is one of the frequently used topics in many applied fields and considered a powerful discriminative method, despite its inability or can not to perform feature selection. The subject of penalized methods is important for the selection of variables. In binary classification with high dimensions, the size of the variables is usually very large, which affects the quality of binary classification. The penalized support vector machine method is considered one of the best methods used because of its performance in selecting features and deferring classification at the same time. The Lp-norm penalty limit is one of the most widely used methods. The panalized support vector machine method depends on appropriately an important value, which is the value of the tuning parameter p. As the choice of this parameter plays an important role in the quality of the penalized support vector machine method. In this thesis, the penalty function Lp was used with the support vector machine and it was proposed to employ the marine predators algorithm for the purpose of reducing the size of the variables and thus obtaining a high classification. For the purpose of verifying the efficiency of the proposed algorithm, its efficiency was examined and compared with the cross validation method (CV) through simulation experiments and applying it to real data that deal with a number of scientific fields. The results show that the superiority or outperformed of the proposed algorithm over the cross-validation method (CV) by selecting the best features and classification accuracy.