

**University of Mosul
College of Computer Sciences
and Mathematics**



Hybrid Swarm Algorithm for Parameters Estimation of the Penalized Support Vector Machine

Niam Abdulmunim Abdulmajeed Al-Thanoon

**Ph. D. Thesis
Statistics**

**Supervised
by**

**Professor
Dr.Zakariya Yahya Algamal**

**Assistant Professor
Dr.Omar Saber Qasim**

1441 A.H.

2019 A.D.

ABSTRACT

High dimensional data has become increasingly common in chemometrics and bioinformatics where frequently, the number of features (variables) surpasses the number of observations. However, the main statistical modeling challenge associated with such data is that the design matrix will not have a full rank. Therefore, the conventional statistical methods, such as regression modeling, are neither applicable nor suitable. Penalized methods have been adapted and have gained popularity as a key for simultaneously performing feature selection and classification. Among the most important classification methods, support vector machine is considered a powerful discriminative method. However, it cannot perform feature selection. Penalized support vector machine (PSVM) proved its effectiveness by creating a strong classifier that combines the advantages of the support vector machine and the penalization. PSVM with L1-norm and smoothly clipped absolute deviation (SCAD) penalty is the most widely used methods. However, the efficiency of PSVM with these penalties depends on appropriately choosing the tuning parameter which is involved in both penalties. In this thesis, swarm algorithms, particle swarm optimization (PSO) and firefly algorithm (FFA) are proposed for effectively handling this issue. Furthermore, a new hybrid firefly algorithm and particle swarm optimization is proposed to determine the tuning parameter in PSVM. Our proposed algorithm can efficiently exploit the strong points of both firefly and particle swarm algorithms in finding the most relevant features with high classification performance. The effectiveness of the proposed methods is examined through simulation studies and applications in the cancer classification and chemometrics real data. The results show that the proposed methods outperformed the cross-validation method (CV) in terms of classification accuracy and the number of selected features.



جامعة الموصل
كلية علوم الحاسوب والرياضيات

خوارزمية السرب المهجنة في تقدير معلمات آلة المتجه الداعم الجزائية

نعم عبد المنعم عبد المجيد الذنون

أطروحة دكتوراه
الأحصاء

بإشراف

الأستاذ المساعد
د. عمر صابر قاسم

الأستاذ الدكتور
زكريا يحيى الجمال

1441 هـ

م 2019

المستخلص

أصبحت البيانات ذات الأبعاد العالية متواجدة ويزيادة في المعلوماتية والحياتية والمعلوماتية الكيميائية والتي فيها يفوق عدد السمات (الصفات، المتغيرات) عدد المشاهدات مما يجعل العديد من الطرائق والأساليب الإحصائية صعبة التطبيق والاستعمال. في السنوات الأخيرة، اكتسبت الطرائق الجزئية شعبية كبيرة بين الإحصائيين كمفتاح لأداء اختيار الصفات (المتغيرات) وتقدير الأنموذج في وقت واحد. يعتبر أسلوب الة المتجه الداعم واحدا من أشهر وأكثر الأساليب استخداما في التصنيف على الرغم من عدم قدرته على أداء اختيار الصفات. يعتبر أسلوب الة المتجه الداعم الجزئي واحدا من أفضل الأساليب المستخدمة بسبب أدائه في اختيار الصفات والارجاء التصنيف في ان واحد. يعتبر حدي الجزاء L1-norm و SCAD من أكثر حدود الجزاء استخداما. يعتمد أسلوب الة المتجه الداعم الجزئي على قيمة مهمة وهي قيمة معلمة الضبط. إذ ان اختيار هذه المعلمة يلعب دورا هاما في جودة أسلوب الة المتجه الداعم الجزئي. تم في هذه الأطروحة اقتراح توظيف خوارزميات السرب مثل خوارزمية سرب الطيور وخوارزمية حشرة اليراعة المضيفة لتقدير معلمة الضبط. علاوة على ذلك تم اقتراح أسلوب هجين جديد بين خوارزمية سرب الطيور وخوارزمية حشرة اليراعة المضيفة. تم اختبار كفاءة الأساليب المقترحة ومقارنتها مع أسلوب العبور المتقاطع CV من خلال اجراء تجارب محاكاة والتطبيق على بيانات حقيقية تختص بالجانب الكيميائي وتصنيف مرضى السرطان من خلال بيانات رقائى الحمض النووي. لقد اظهرت النتائج تفوق الأساليب المقترحة على أسلوب CV من خلال اختيار افضل السمات وباعلى تصنيف. كذلك اظهرت النتائج تفوق أسلوب التهجين الجديد المقترح على كل من أسلوب خوارزمية سرب الطيور واسلوب خوارزمية حشرة اليراعة المضيفة.