ทำความเข้าใจเกี่ยวกับ SVM

เป้าหมาย

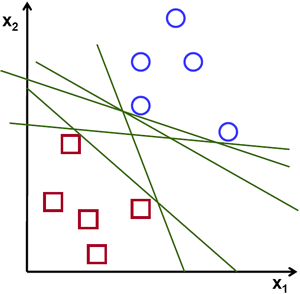
ในบทนี้

* เราจะเห็นความเข้าใจง่ายของ SVM

ทฤษฎี

ข้อมูลที่แยกได้โดย Linear

พิจารณาภาพด้านล่างซึ่งมีข้อมูลสองประเภทคือสีแดงและสีน้ำเงิน ใน kNN สำหรับข้อมูลการทดสอบเราใช้ในการวัดระยะทางของมันไปยังตัวอย่างการฝึกอบรมทั้งหมดและใช้ระยะทางอย่างน้อยที่สุด ใช้เวลาในการวัดระยะทางและหน่วยความจำมากมายเพื่อเก็บตัวอย่างการฝึกอบรมทั้งหมด แต่ถ้าพิจารณาจากข้อมูลที่ให้ไว้ในภาพเราควรต้องการอะไรมากหรือไม่?



พิจารณาความคิดอื่น เราพบบรรทัดf (x) = ax_1 + bx_2 + Cซึ่งแบ่งข้อมูลทั้งสองไปยังภูมิภาค เมื่อเราได้รับ test\_data ใหม่เพียงแค่ทดแทนในX f (x)ถ้าf (X)> 0มันเป็นของกลุ่มสีฟ้าอื่นมันเป็นของกลุ่มสีแดง เราสามารถเรียกสายนี้เป็นเขตแดนตัดสินใจ มันง่ายมากและมีหน่วยความจำที่มีประสิทธิภาพ ข้อมูลดังกล่าวซึ่งสามารถแบ่งออกเป็นสองด้วยเส้นตรง (หรือ hyperplanes ในมิติที่สูงขึ้น) เรียกว่าเป็น Linear แบบแยก

ดังนั้นในภาพด้านบนคุณจะเห็นเส้นดังกล่าวเป็นไปได้ ที่เราจะใช้? มากอย่างสังหรณ์ใจเราสามารถพูดได้ว่าสายควรจะผ่านเท่าที่เป็นไปได้จากทุกจุด ทำไม? เนื่องจากอาจมีเสียงรบกวนในข้อมูลที่เข้ามา ข้อมูลนี้ไม่ควรมีผลต่อความถูกต้องในการจัดหมวดหมู่ ดังนั้นการใช้สายมากที่สุดจะทำให้ภูมิคุ้มกันลดลง ดังนั้นสิ่งที่ SVM ทำคือการหาเส้นตรง (หรือไฮเปอร์เพลน) ที่มีระยะห่างขั้นต่ำที่สุดไปยังตัวอย่างการฝึกอบรม ดูเส้นหนาด้านล่างภาพที่ผ่านใจกลาง



เพื่อหาขอบเขตการตัดสินใจนี้คุณต้องมีข้อมูลการฝึกอบรม คุณต้องการทั้งหมดหรือไม่? NO เพียงแค่คนที่อยู่ใกล้กับกลุ่มตรงข้ามก็เพียงพอแล้ว ในภาพของเราพวกเขาเป็นวงกลมที่เต็มไปด้วยสีน้ำเงินและสองช่องสีแดงเต็มไปหมด เราสามารถเรียกพวกเขาเวกเตอร์การสนับสนุนและสายผ่านพวกเขาจะเรียกว่าเครื่องบินสนับสนุน พวกเขามีความเพียงพอในการหาขอบเขตการตัดสินใจของเรา เราไม่จำเป็นต้องกังวลเกี่ยวกับข้อมูลทั้งหมด ช่วยในการลดข้อมูล

สิ่งที่เกิดขึ้นคือพบว่ามีสอง hyperplanes แรกที่แสดงข้อมูลได้ดีที่สุด สำหรับเช่นข้อมูลสีฟ้าเป็นตัวแทนจากw ^ Tx + b_0> 1ในขณะที่ข้อมูลสีแดงเป็นตัวแทนจากw ^ Tx + b_0 <-1ที่Wเป็นน้ำหนักเวกเตอร์ ( w = [w_1, w_2, ... , w_n]) และxเป็นเวกเตอร์คุณลักษณะ ( x = [x_1, x_2, ... , x_n]) b_0เป็นอคติ เวกเตอร์น้ำหนักตัดสินใจแนวการกำหนดขอบเขตการตัดสินใจในขณะที่จุดอคติจะตัดสินใจที่ตั้ง ตอนนี้ขอบเขตการตัดสินใจถูกกำหนดให้เป็นกึ่งกลางระหว่าง hyperplanes w ^ Tx + b_0 = 0เหล่านี้เพื่อแสดงเป็น ระยะห่างต่ำสุดจากเวกเตอร์สนับสนุนไปยังขอบเขตการตัดสินใจจะกำหนดโดย, distance_ {support \, vectors} = \ frac {1} {|| w ||}. Margin เป็นสองเท่าของระยะนี้และเราต้องเพิ่ม margin นี้ เช่นเราจำเป็นต้องลดฟังก์ชันใหม่L (w, b_0)ด้วยข้อ จำกัด บางอย่างที่สามารถแสดงด้านล่างได้:

\ min_ {w, b_0} L (w, b_0) = \ frac {1} {2} || w || ^ 2 \;  \ text {subject to \;  t_i (w ^ Tx + b_0) \ geq 1 \;  \ forall i

ที่เป็นฉลากของแต่ละชั้นเรียนTit_i \ in [-1,1]

ข้อมูลที่แยกได้ไม่เป็น Linearly

พิจารณาข้อมูลบางอย่างที่ไม่สามารถแบ่งออกเป็นสองเส้น ตัวอย่างเช่นพิจารณาข้อมูลแบบ 1 มิติที่ 'X' อยู่ที่ -3 และ +3 และ 'O' อยู่ที่ -1 และ +1 เห็นได้ชัดว่ามันไม่สามารถแยกได้เป็นเส้นตรง แต่มีวิธีการแก้ปัญหาแบบนี้ ถ้าเราสามารถแม็พชุดข้อมูลนี้ด้วยฟังก์ชันf (x) = x ^ 2เราจะได้ค่า 'X' ที่ 9 และ 'O' ที่ 1 ซึ่งเป็นเส้นที่แยกออกได้

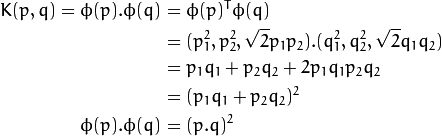
มิฉะนั้นเราสามารถแปลงข้อมูลแบบสองมิตินี้เป็นมิติเดียว เราสามารถใช้f (x) = (x, x ^ 2)ฟังก์ชันเพื่อทำแผนที่ข้อมูลนี้ได้ จากนั้นค่า "X" จะกลายเป็น (-3,9) และ (3,9) ขณะที่ "O" จะกลายเป็น (-1,1) และ (1,1) นี้ยังเป็นเส้นที่แยกออกได้ ในระยะสั้นโอกาสสำหรับข้อมูลที่ไม่สามารถแยกได้ในเชิงเส้นในพื้นที่ที่มีขนาดต่ำกว่าจะกลายเป็นเส้นที่แยกได้ในพื้นที่ที่มีมิติสูงกว่า

โดยทั่วไปมันเป็นไปได้ที่จะทำแผนที่จุดในพื้นที่มิติ d- มิติบางมิติ(D> ง)เพื่อตรวจสอบความเป็นไปได้ที่จะแยกได้ มีแนวคิดที่จะช่วยในการคำนวณจุดผลิตภัณฑ์ในพื้นที่ขนาดใหญ่ (kernel) ด้วยการคำนวณในพื้นที่ป้อนข้อมูล (มิติ) ขนาดต่ำ เราสามารถอธิบายด้วยตัวอย่างต่อไปนี้

พิจารณาสองจุดในพื้นที่สองมิติและp = (P_1, P_2) q = (q_1, q_2)อนุญาต\ พีเป็นฟังก์ชันการทำแผนที่ที่แม็ปสองมิติกับพื้นที่สามมิติดังต่อไปนี้:

\ phi (p) = (p_ {1} ^ 2, p_ {2} ^ 2, \ sqrt {2} p_1 p_2) \ phi (q) = (q_ {1} ^ 2, q_ {2} ^ 2, \ sqrt {2} q_1 q_2)

ให้เรากำหนดฟังก์ชั่นเคอร์เนลK (P, Q)ที่ทำผลิตภัณฑ์จุดระหว่างสองจุดที่แสดงด้านล่าง:

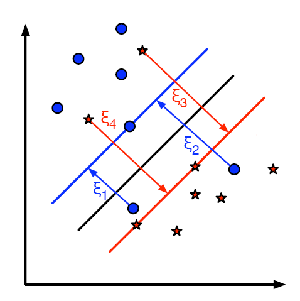


หมายถึงผลิตภัณฑ์จุดในพื้นที่สามมิติสามารถทำได้โดยใช้ผลิตภัณฑ์สแควร์ในพื้นที่สองมิติ นี้สามารถใช้กับพื้นที่มิติที่สูงขึ้น ดังนั้นเราจึงสามารถคำนวณคุณสมบัติมิติสูงขึ้นจากมิติข้อมูลที่ต่ำกว่าได้ เมื่อเราทำแผนที่แล้วเราจะได้พื้นที่ในมิติที่สูงขึ้น

นอกเหนือจากแนวคิดเหล่านี้แล้วยังมีปัญหาเรื่องการจัดประเภทที่ไม่ถูกต้อง ดังนั้นการหาขอบเขตการตัดสินใจที่มีอัตรากำไรสูงสุดไม่เพียงพอ เรายังต้องพิจารณาถึงปัญหาของข้อผิดพลาดในการจัดหมวดหมู่ด้วย บางครั้งอาจเป็นไปได้ที่จะหาเขตแดนการตัดสินใจที่มีอัตรากำไรน้อย แต่มีการจัดหมวดหมู่ที่ลดลง อย่างไรก็ตามเราจำเป็นต้องปรับเปลี่ยนรูปแบบของเราเพื่อที่จะหาขอบเขตการตัดสินใจที่มีอัตรากำไรสูงสุด แต่มีการจัดระดับไม่ถูกต้อง เกณฑ์การปรับลดถูกปรับเปลี่ยนเป็น:

min \;  || w || ^ 2 + C (ระยะทาง \; ของ \; ที่ถูกจัดประเภทอย่าง \; ตัวอย่าง \ ไปยัง \; ถูกต้อง \; ภูมิภาค)

ด้านล่างภาพแสดงแนวคิดนี้ สำหรับตัวอย่างของข้อมูลการฝึกอบรมแต่ละตัว\ xi_iจะมีการกำหนดพารามิเตอร์ใหม่ เป็นระยะห่างจากตัวอย่างการฝึกอบรมที่สอดคล้องกันในขอบเขตการตัดสินใจที่ถูกต้อง สำหรับผู้ที่ไม่เข้าใจผิดพวกเขาตกบนเครื่องบินสนับสนุนของพวกเขาดังนั้นระยะทางของพวกเขาเป็นศูนย์



ดังนั้นปัญหาการเพิ่มประสิทธิภาพใหม่คือ:

\ min_ {w, b_ {0}} L (w, b_0) = || w || ^ {2} + C \ sum_ {i} {\ xi_ {i}} \ text {subject to} y_ {i} (w ^ {t} x_ {i} + b_ {0}} \ geq 1 - \ xi_ {i} \ text {and} \ xi_ {i} \ geq 0 \ text {} \ forall i

ควรเลือกพารามิเตอร์ C อย่างไร? เป็นที่ชัดเจนว่าคำตอบสำหรับคำถามนี้ขึ้นอยู่กับข้อมูลการฝึกอบรมที่กระจายอย่างไร แม้ว่าจะไม่มีคำตอบทั่วไป แต่ก็ควรคำนึงถึงกฎเหล่านี้ด้วย:

ค่าขนาดใหญ่ของ C ให้โซลูชันที่มีข้อผิดพลาดในการจัดระดับที่น้อยกว่า แต่มีขนาดเล็กลง พิจารณาว่าในกรณีนี้จะมีข้อผิดพลาดในการจัดชั้นไม่ดี เนื่องจากจุดมุ่งหมายของการเพิ่มประสิทธิภาพคือการลดอาร์กิวเมนต์ให้อนุญาตข้อผิดพลาดในการรวบรวมข้อมูลผิดพลาดเพียงไม่กี่รายการ

ค่าขนาดเล็กของ C ให้โซลูชันที่มีอัตรากำไรสูงกว่าและมีข้อผิดพลาดในการจัดหมวดหมู่มากขึ้น ในกรณีนี้การลดลงไม่ได้พิจารณาว่าระยะเวลาของการรวมดังนั้นจึงมุ่งเน้นเพิ่มเติมเกี่ยวกับการหาไฮเปอร์เพลนที่มีอัตรากำไรสูง