Açıklama Üretim sektöründe aktif faaliyet gösteren bir şirkette çalıştığınızı farzediniz. Üst yönetim üretim hattının verimini arttırmak ve olası robot arızalarını önceden tahminlemek istiyor. Üretim hattında 3 robotun olduğunu düşünelim. A robotu → B robotu → C robotu Robotlar sırasıyla ürünü alır, işlemini yapar ve ürünü bırakır.

Sorular

**1** Üretim hattının verimini arttırmak ve olası robot arızalarını önceden tahminleyebilmek için nasıl bir verisetine ihtiyacınız olurdu?

* İstediğiniz verisetinde bulunan verilere/kolonlara neden ihtiyaç duyduğunuzu açıklanmanız ve birkaç satırlık veriseti örneği eklemeniz beklenmektedir.

**2** Üretim hattının verimini arttırmak ve olası robot arızalarını önceden tahminleyebilmek için nasıl bir algoritma kurardınız?

* Algoritmanın adımlarını ve adımların açıklamalarını yazmanız beklenmektedir.

**3** Kurduğunuz algoritmadan çıkan sonuçların üst yönetime real-time görselleştirilmesi için nasıl bir teknolojik mimari tasarım oluştururdunuz?

* Herhangi bir diyagram çizme uygulaması (ÖRN:LucidChart) kullanarak tasarımınızı çizmeniz ve adımları açıklamanız beklenmektedir

**1.Soru**

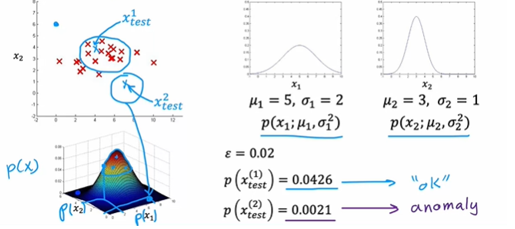
Olası robot arızalarını önceden tahminleyebilmek için veriseti denetimli öğrenmediki(supervised learning) gibi etiketli olamaz çünkü bilinmeyen hatalar bütünü bu yüzden denetimsiz öğrenme(unsupervised learning) yolu kullanılmalı, gelecekte olası hataları kestirmek ve engellemek için anolmaly detection en etkili yoldur. Anormallik algılama algoritmaları, normal olayların etiketlenmemiş bir veri kümesine bakar ve böylece olağandışı veya anormal bir olay olup olmadığını algılamayı tespit etmeyi öğrenir. Robot örneğinde de mesela x1 özelliği sıcaklığı ve x2 özelliği ise titreşim yoğunlu olsun.

Öğrenme algoritması, robotların ne kadar ısı üretildiği ve ne kadar titreştikleri açısından tipik olarak nasıl davrandığına dair geçmişten m örneklerinin elimizde olsun.

Anomali tespitini gerçekleştirmenin en yaygın yolu, yoğunluk tahmini adı verilen bir tekniktir. Bunun anlamı, m örnekten oluşan eğitim kümeleri olduğundan, yapılan ilk şey, x'in olasılığı için bir model oluşturmaktır. Başka bir deyişle, öğrenme algoritması, x1 ve x2 özelliklerinin yüksek olasılığa sahip değerlerinin neler olduğunu ve veri setinde görülme olasılığı daha büyük veya daha düşük olan değerlerin neler olduğunu bulmaya çalışacaktır. Bu x1 ve x2 özelliği örnek olarak verilmiştir çok daha fazla örnek ekleyip olasılık yoğunluk fonksiyonları oluşturulabilir.

x1'den xm'ye kadar bir eğitim seti var, burada her x örneğinin bitiş özellikleri var. Yani, her x örneği sayılar içinde bir vektördür. Robot örneğinde, ısı ve titreşimlere karşılık gelen iki özelliğimiz vardır.

**2.Soru**

Anormallik tespitini uygulamak için normal dağılım olarak da adlandırılan Gauss dağılımın kullanılması gerekir. yapılmak istenen yoğunluk tahminidir ve bunun tek anlamı, bir model kurulacak ve p(x) için olasılığı tahmin edilecek. p(x) için modelimiz şu şekilde olacaktır, x, x1, x2 ve xn'ye kadar devam eden değerlere sahip bir özellik vektörüdür. Her değerin olasılığı hesaplanır, her bir değerin olma olasılığı istatisksel olarak birbirinden bağımsızdır. Aslında algoritma mantığı olasılık yoğunluk fonksiyonlarının dağılımına göre olası hataları tespit etme üzerine kuruludur. Mesela x1 ve x2 için olasılığını bilip anolmaly olup olmadığını şöyle anlayabiliriz;

Olası robot arızasını tespit sistemini nasıl oluşturabileceğinizi görmek için hepsini bir araya getirilmesi gerekir. İlk adım, anormal örneklerin göstergesi olabileceğini düşündüğünüz özellikleri (xi) seçmek olacaktır; yukarıdaki p(x2) değeri gibi değerler. Bu sayede etiketlenmemiş eğitim setindeki bu parametreleri tahmin ederek, artık modelimizin tüm parametrelerini hesaplamış oluyoruz. Bir olasılık sonucunun anomaly olup olmadığını threshold ile seçebiliriz.

Anomaly tespitinin iyi çalışıp çalışmadığını Threshold değeri belirler, bu değerin seçimi önemlidir. Algoritmayı bir özelliği değiştirmek veya bir parametreyi değiştirmek gibi bir şekilde hızlı bir şekilde değiştirebilirsek ve algoritmanın daha iyi veya daha kötü olup olmadığını bize söyleyen bir sayı hesaplama yöntemine sahipsek(cross validation test), o zaman algoritmanın daha iyi veya daha kötü olup olmadığına karar vermeyi çok daha kolaylaştırır. Bu değişikliğe algoritmaya bağlı kalmak için anomali tespitinde genellikle bu şekilde yapılır. Anolmaly tespit sisteminin geliştirilmesi için birkaç adım da uygulanabilir mesela belki birkaç robot çalışmasını yaptıktan sonra, anormal olan birkaç robot görüldü ve anormal olduğunu bildiğiniz örnekler için, bu anormalliği belirtmek için y eşittir 1 etiketini ilişkilendirelim ve normal olduğunu düşündüğümüz örnekler için ise y etiketi 0'a eşit olur.

Anomali tespit algoritmasının öğreneceği eğitim seti hala x1'den xm'e kadar olan bu etiketsiz eğitim setidir ve tüm bu örnekleri sadece normal ve anormal olmadığını varsayacağımız örnekler olarak düşünebiliriz, bu yüzden y eşittir 0. Geçmişteki yapılan robotlardan elde edilen anormal örneği varsa, çapraz doğrulama kümesi oluşturabilmek için çok yararlı olacaktır, bu sayede hem çapraz doğrulama hem de test setlerinin birkaç anormal örnek içerdiği bazı örneklerden oluşan bir test setine sahip olmuş oluyoruz. Özetle anormal olan bazı örnekler varsa, ancak yanlışlıkla y eşittir 0 ile etiketlenmişse, anormallik saptama algoritması iyi çalışacaktır.

Örneklemek gerekirse Threshold değerini seçmek için denetimli öğrenme(supervised learning) kullanılır bu threshold değerini ise denetimsiz öğrenme(unsupervised learning) için kullanılır. Threshold değerinin önceki elde edilen etiketlemiş anolmaly ve normal verilerle çapraz doğrulama testi ile deneme yanılma yoluyla seçebiliriz, anolmaly bildiğimiz etiketleri bulduğu zaman threshold parametresini bulmul oluruz daha sonra ise gauss dağılımları ile etiketsiz veriler üzerinde robotun iyi çalışıp çalışmadığını ısı ve titreşim yoğunluk fonksiyonlarına bakarak olası hataları tespit ederiz, biraz karışık gözüküyor ama basitçe tekrar etmek gerekirse; ısı ve titreşim verilerinden oluşan bir etiketsiz veri setimiz var ancak bunlar arasında korelasyonda neye göre anolmaly veya normal olacağını kestirmek için threshold değerini bilemiyoruz, bu değeri bilmek için ise geçmişteki anomaly ve normal etiketli veriler ile çarpraz doğrulma testi ile threshold değerini bilebiliyoruz bu sayede gelecekte farklı hiç görülmemiş değerler ile olasılık yoğunluk fonksiyonları oluşturup threshold ile anolmaly tespiti yapıp **olası robot arızalarını önceden tahminleyebiliriz.**

Burada dikkat çekmek istediğim nokta robot arızasının tespitini denetimli öğrenme ile yapabiliriz ancak **olası hatalar** için geçmişteki hataların bilinmesi gelecekteki hataları kesin olarak vermeyecektir bu yüzden anomaly detection kullanılacaktır.

Bir anormallik tespit algoritması oluştururken, iyi bir özellik seçimi seçmenin(bizim varsayımız da sıcaklık ve titreşimdi) önemlidir. Denetimli öğrenmede, özelliklere tam olarak sahip değilsek veya sorunla alakalı olmayan birkaç ekstra özellik varsa, bu genellikle iyi olur. Algoritmanın denetimli sinyal vermesi gerektiğinden, algoritmanın hangi özelliklerin yok sayıldığını veya özelliğin nasıl yeniden ölçeklendirileceğini ve ona verilen özelliklerden en iyi şekilde yararlanabileceğini anlaması için yeterli etiket vardır. Ancak, yalnızca etiketlenmemiş verilerden çalışan veya öğrenen anomaly tespiti için, anomalinin hangi özelliklerin göz ardı edileceğini anlaması daha zordur. Bu yüzden, özellikleri dikkatlice seçmenin, anormallik tespiti için denetimli öğrenme yaklaşımlarından daha önemlidir.

Seçilen özellik etiketsiz verisi normal dağılım veya sağ\sol yanlı gauss olmalıdır, yanlı gauss verileri matematiksel denklemlerle normal dağılıma çekilebilir.

