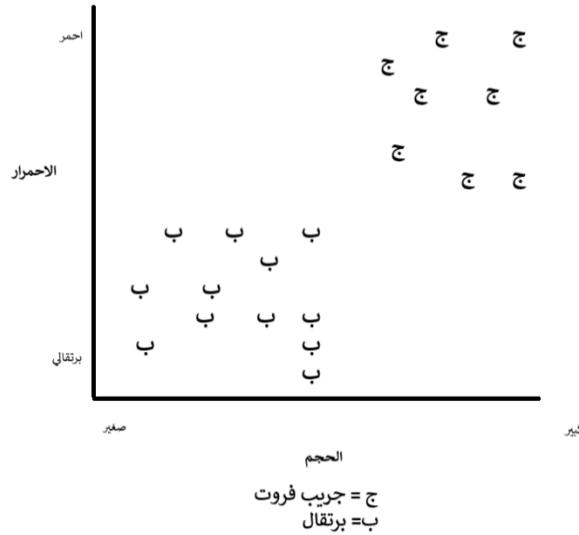


. في الجزء ده هنعرف ازاى ممكن نعمل خوارزمية تصنيف وشوية مفاهيم في تعليم الآلة (machine learning)

## البرتقال والجريب فروت

لو انا حظيت قدامك فاكهة وسألتك هو ده برتقال ولا جريب فروت فا انت غالبا هتبدأ تفكر ف بالخصائص الي تميز الجريب فروت عن البرتقال وتشوف الفاكهة ال قدامك دي بينتبق عليها انهى صفات من الي انت حددتها في الآخر غالبا هتوصل لان الجريب فروت أكبر ولونها احمر اكثر من البرتقال ودي حاجة ممكن نمثلها بالرسم ده



الرسمه دي بتوضح الفكرة اكثر بس في مشكلة صغيرة في مكان فاضي في النص ده منطقة بتمثل بالنسبالي شك بمعنى ان الفاكهة الي مواصفاتها هتخليها فالمكان ده ان امش هقدر اجزم بنوعها طب هنتصرف ازاى هنشوف أقرب جيران الفاكهة بتاعتنا لو عدد الجيران الي كانوا جريب فروت كانوا اكثر يبقى هي جريب فروت والعكس الخوارزمية دي اسمها KNN ودي خوارزمية تصنيف خلونا نشوف مثال ثاني

## نظام الترشيح

دلوقتي نفترض أنك نيتفليكس وعايز تعمل نظام ترشيحات يرشح للمشاهدين أفلام يتفرجوا عليها زي ما عملنا في المثال الي فات هنحط كل المستخدمين الي عندنا في رسم بياني فا المستخدمين الي شبه بعض هيكونوا قريبين من بعض وبكده لو واحد من المستخدمين عجبه فيلم القيم ده هيترشح للمستخدمين الي شبهه

## استخراج الخصائص

احنا دلوقتي ناقشنا مثالين مختلفين و احنا نقدر بنفسنا نعرف النقط المتشابهة بنفسنا بسهولة لكن الكمبيوتر بالنسبالي الموضوع مش بالسهولة دي يعني في مثال الجريب فروت احنا نقدر نقول اذا كانت الفاكهة لونها احمر ولا لا و مدى الاحمرار ده لكن الكمبيوتر ميقدرش فا احنا هنحط مقياس من 1 ل 5 بحيث ان 1 يكون برتقالي و يزيد الاحمرار لحد ما توصل ل 5 و نف سالكلام مع الحجم بالطريقة دي نقدر نعمل الرسم البياني بسهولة بشكل الكمبيوتر يفهموا وبرضه لازم تخلي في دماغك انك تختار خصائص تساعدك يعني مثلا في مثال الجريب فروت لو انا اخترت اني أقيم الفاكهة على أساس هي مدورة ولا لا فا دي حاجة ملهاش لازمة عشان الاتنين مدورين وده ممكن يأدي لتصنيف غلط في الآخر طب انا في مثال نظام الترشيحات هختار الخصائص بتاعتي على أساس ايه لما بتعمل حساب على نيتفليكس هيطلب منك تقييم شوية تصميقات الكوميديا, الاكشن, الدراما, الرعب والرومانسية على الأساس بيتعمل الرسم البياني بس لو لاحظت النقطة هنا ليها خمس ابعاد و احنا

منعرفش ازاي نرسم خمس ابعاد فا احنا في الحقيقة مينرسمش فاكر المعادلة الي بتحسب المسافة بين نقطتين  $v((x_2 - x_1)^2 + (y_2 - y_1)^2)$  المعادلة دي هتشتغل معنا مهما كان عدد الابعاد بتاعتي حتى لو كانت مليون و بالمعادلة دي اقدر اعرف مين اقرب جيرانني لو قلنا مثلا ان عندنا اتنين مشتركين في نيتفليكس النقط الي بتمثلهم هي دي  $(3,2,1,5,4)$  و  $(5,2,1,3,0)$  المعادلة ده هيكون شكل المعادلة  $v((5-3)^2 + (2-2)^2 + (1-1)^2 + (3-5)^2 + (0-4)^2)$  و الي هنطلعنا ف يالاخر ان المسافة هي 2.76

## الانحدار (regression)

مش دايمًا هيكون هدفي أني اصنف الحاجات الي قدامي ساعات هكون عايز أتوقع يعني مثلاً بعد ما أرشح فيلم لحد انا عايز اعرف هو هيديله تقييم كام فمثلاً لو انا رشتلك فيلم من اقرب خمس جيران ليك و كانت تقييماتهم على مقياس من 1 ل 5 كده 5,4,5,4,3 تقدر تتوقع التقييم عن طريق انك تحسب متوسط التقييمات دي الي في الحالة دي هيكون 4.2 هو ده الانحدار

ملحوظة: عدد الجران الي انت بتخترهم ممكن يختلف عادي بس لازم تاخذ بالك انه ميكنش قليل جدا او كبير جدا عشان ممكن يادي لتصنيفات غلط

## مقدمة في تعليم الآلة

تعليم الآلة هو أنك تخلي الكمبيوتر بتاعك أكثر ذكاء انت شفت مثال على ده و احنا بنستخدم KNN خليةنا نشوف شوية امثلة كمان

**OCR:** هي الطريقة الي بنخلي الكمبيوتر يتعرف بيها على الحروف والأرقام لما تكون موجودة في صور يعني انت ممكن تاخذ صورة لصفحة من كتاب و الكمبيوتر هيعرف يقرأها ويفهم ايه الحروف والكلمات الموجودة في الصورة و ده جوجل مثلاً بتستخدمه عشان تعمل نسخ رقمية من الكتب طب ازاي ده بيحصل؟

خلينا نقول دلوقتي ان معانا صورة للرقم 7 عشان أخلي الكمبيوتر يتعرف عليها هعمل الاتي  
1- هخلي الكمبيوتر يشوف صور مختلفة للأرقام ويتعرف على الخصائص بتاعتها (هنا انا بعلم الكمبيوتر الخصائص بتاعت كل رقم انا بعرفه ازاي بيكون شكل رقم 7 وازاي بيكون شكل رقم 3 وهكذا)  
2- لما الكمبيوتر هيجيله صورة جديدة هيشوف الخصائص بتاعتها ويقارنها بجوانها وعلى أساس ده هيعرف ايه الرمز الي في الصورة زي مثال الجريب فروت

ملحوظة: في حالة الحروف والأرقام الخصائص بالنسبة الي بتكون الخطوط والنقط والانحناءات

## Spam filter

دلوقتي لو انا عندي مجموعة من الرسائل وعايز اعرف اذا كانت فعلاً رسالة حقيقية ولا في حد بيحاول ينصب عليا الموضوع ده بيتعمل بخوارزمية لسمها naïve bayes اول حاجة بتحتاج انك تدرب البرنامج بتاعك

مجموعة التدريب

الموضوع	سبام؟
"غير كلمة المرور بتاعتك"	مش سبام
"انت كسبت مليون دولار"	سبام
"ابعتلي كلمة المرور بتاعتك"	سبام
"امير نيجيري باعتلك 10 مليون دولار"	سبام
"كل سنة وانت طيب"	مش سبام

دلوقتي جالك رسالة جديدة بتقول "استلم المليون دولار بتوعك" الخوارزمية بتاعتنا هنا هتقسم الرسالة لكلمات وبتشوف احتمالية وجود كل كلمة في رسالة سبام في المثال البسيط بتاعنا ديما كلمة كليون بتكون موجودة في رسائل سبام فهو هيعتبر الرسالة دي سبام.

## توقع أسعار الاسهم

هنا دي حاجة صعب نعلم الكمبيوتر انه يعملها لان الموضوع بيعتمد على متغيرات كثيرة ومفيهوش ثوابت مقدرش أقول مثلا ان بما ان السعر السهم ارتفع امبارح فهو هيرتفع بكرة او ان سعر السهم ده بيرتفع ديما في شهر خمسة معنديش طريقة واضحة أقدر استخدم فيها الماضي عشان أتوقع المستقبل فالموضوع يعتبر شبه مستحيل.

## النهاية

دي كانت كل حاجة معانا في الجزء ده وفي الكتاب كله الملخص ده ممكن سكون مفيد ليك لو انت قرئت الكتاب وبتراجع او بتتعرف على محتوى الكتاب بشكل عام فا أتمنى انه يكون ساعدك بأي شكل من الاشكال و أخيرا في باب كمان موجود فالكتاب بس هو عبارة عن أفكار لمواضيع ممكن تبدأ تتعلمها بعد ما تخلص الكتاب فا انا شفت انه يفضل تقراه من الكتاب.