

Statistics and Probability



=> في ال statistics بنقيس 3 حاجات :

1. Central Tendency (Mean, Median, Mode, Trimmed mean)
2. Distribution and density
3. Measure of variance

=> ال Central tendency او النزعه المركزيه معناها الداتا متركزة حولين اني نقطة , طيب انا هستفيد ايه لما اجيب قيمتها ؟

=> هستفيد من ده بأني اعرف ايه الموصفات الي تنفع مع اكبر قدر من الداتا ودايما بنميل اننا نجيب القيمه المتوسطه عشان نعمل حسابتنا عليها

=> في Cases بنحسب فيها ال Warst Case لأنها لو حصلت هتبقى كارته كبيره

=> وفي Cases ثانيه لا مش مضطرين اننا نحسبها وبنحسب ال Central Tendency لأن لو ال Warst case حصلت في الحالات دي مش هتبقى كارته +

لأن حساب ال Warst case بيكلفني وقت ومجهود "مثال التمويل والأسر الي اعدد الاطفال فيها من 0 ل 15 طفل "

=< كل اما ال Mean وال Median يكونوا بُعاد عن بعض كل اما الداتا تكون unbalanced

=< ال Trimmed Mean هي اني ارتب الداتا واقص اكبر قيم واصغر قيم بس بشرط اني مقصش اكثر من 10% من الداتا

=< انت معاك داتا كثير وبتحاول تجيب ال Central tendency بتاعتها بأما هتضطر تاخذ ال mean or median or mode or trimmed mean اي واحد فيهم بس تكون بتعبر عن الداتا بشكل كويس وده عشان احقق ال Concept ال generlaization الي هو الغرض منه اننا ن optimize

=< طب افرض حسبنا ده كله وملقناش ولا قيمه بتعبر عن الداتا ؟ ده معناه ان الداتا هي الي فيها مشكله ومحتاجين اننا نغيرها

=< ال Distribution and Density :

=< Based على النزعه المركزيه ممكن اخذ قرار وممكن حاسبها ميكونش كافي اني اخذ قرار فهنا في حاجه اسمها التوزيع والكثافه

=< في فرق ما بين التوزيع والكثافه :

- خلونا ناخذ مثال وهو مثال التيشترات بتاعت حد مشهور مثلا نزل السوق المصنع الي نزل التيشترات الناس الي اشترتها كانت الي مقاستهم XL, M, L بس فده كده اسمه التوزيع او **Distribution** لما جم يشوفوا نسب الناس الي اشترت في كل مقاس فيهم لقوا ان اكثر ناس اشترت هما الي مقاستهم L فده ال **Density**
- يبقى فكرة التوزيع ان انا معايا Range فبشوف الداتا متوزعه في اني اطار في ال Range
- فالاستفاده من ال **Distribution and Density** انه اعرف الداتا الأكثر انتشاراً في Range قد ايه او بتقع في اني Range

- وكمان اشوف اكثر داتا موجودة فين في ال Range ده بالظبط عشان اشوف ايه اكثر داتا هكون مهتمه بيه

=< ال data analysis ما هي ال Statistics

=< ال curtails وال percentiles؟

- ال curtails هو اني بقسم الداتا بتاعتي 4 اربع
- ال percentiles هو اني بحط ال sections الي انا بوزع على اساسها على هيئة نسب مئوية بهدف اني اتعامل مع كل نسبه او كل كميه بطريقة معينه
- وبستخدمهم في حالة ان الحل الي احنا عاوزين ن generalize على الداتا مينفعش لكل الانواع فبالتالي بنقسم الداتا ل groups معينه

=< كل 3 شهور في ال business عبارته عن حاجه اسمها cortal

- لأنه بيقلولي ان اكثر داتا بتكون موجوده في المنتصف يعني ال distribution بيكون normal لأن اكبر نسبه من الداتا بتكون في المنتصف والداتا الي القيم بتاعتها كبيره اوي او صغيره اوي فنسبتهم قليله جدا + ان دي الفطره الي موجوده في الكون وربنا خلق الكون عليها
- كان زمان معظم الداتا كانت normal

=< هل ينفع اعمل analysis على داتا مش في ال normal distribution ؟

- اه ممكن اعمل analysis عليها بنفس شكلها
- او ممكن احوّلها لل normal distribution

=< طيب ليه نحاولها ل normal ؟

- ده عشان النورمال هو الأساس
- ولأن لو انت عملت analysis على normal distribution فالناتج الي هيطلع انت هتكون متأكد منه بنسبه كبيره انه صح
- النورمال معناه ان الداتا بتاعتي balanced يعني ال mean وال median هيكونوا منطبقين على بعض يعني نسبة كبيره من القرار الي انا هاخده based على شكل الداتا دي هيكون صح

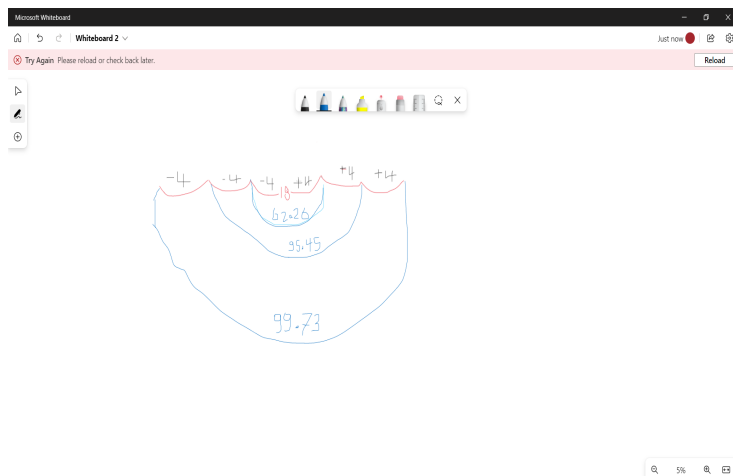
=< ال Statistics الهدف منها هو عملية ال decision making

=< ال Measure of variance هو اننا بنقيس نسب الاختلاف في الداتا طيب
هنستفيد ايه من ده ؟

- كل اما تكون الداتا متشبهها مع بعض كل اما طرق تعاملتي مع الداتا هتقل بمعنى انه لو معظم الداتا متشابهه فانا هتعامل معاها بطريقة واحده
- يبقى ال variance كل اما يكون عالي كل اما هتضطر اني اتعامل مع الداتا بشكل مختلف " زي مثال شركات الاتصالات وانواع العمله الي بتتكلم "
- اخر مره كان ال varaince فيها قليل كان قبل ظهور الانترنت
- كل اما زاد ال varaince كل اما زاد احتمالات القرارات الكيتر الي ممكن اخدها
- بنقيسه عن طريق اننا نجيب بُعد كل نقطه عن المتوسط
- ممكن ال variance يكون قليل بس كبير في نفس الوقت "مثال تحليل السكر مثلا لو ال variance مثلا 0.5 ممكن يقول انه الراجل ده عنده سكر وهو مكنش عنده مثلا "

=< ايه هو ال standard deviation ؟

- لما يكون عندنا داتا زي دي مثلا 10,15, 20,17, 22,30 وجيت قولتلك ان ال mean =18 , std =4 فده معناه ايه ؟



- ده معناه ان اغلب الداتا بتقع عند المتوسط لحد بعديه ب 4 و قبله ب 4 بمعنى mean+std , mean-std4
- النسب دي معناها ان 68.26% من الداتا بيقع في ال std ده
- والي بيقع خارج النسب دول بيقي outlier

=< ال Z-Score :

- لو انا معايا داتا عباره عن شوية ارقام وال mean =20 وال Std =5 وجيت قولتلك ايه هو ال z-Score بتاع احد عناصر الداتا الي اسمها 15

- ده معناه ان ال 15 دي بعيده عن ال mean بكام std يعني القانون هنا
- $X\text{-Mean} / \text{STD} = 1$ هنا مثلا بالتعويض هيطلع -1 فده معناه ان ال 15 بعيده عن ال Mean ب1 std من قبله او من شمال ال mean
- طيب ex على ده في الواقع زي مثلا محتاجين نعمل اكتوبر للطلاب يقولنا هنعط الطالب ده في فصل المتفوقين ولا المتوسطين ولا الضعاف

=< ايه هي ال Hypothesis testing :

- هو اني عندي نظريه بحاول اثبت ان عكسها صح لو عكسها صح فده معناه النظرية المفروضه غلط ولو عكسها مش صح فده مش معناه ان النظرية صح هي صح لحد ما يتم اثبات العكس
- الافتراضيه اسمها Null Hypothesis , وعكس الافتراضيه اسمها alternative hypothesis
- اثبات ان النظرية نفسها صح او غلط اصعب بكثير من اننا نشتغل على عكسها " المثال ان الساعه جات 10 ونص وانا لسه مجتش "وده مش هيحصل عشان انا حد جامد جدا ربنا يحفظني "

=< ال p-Value :

- هو نفس المثال بتاع اني مش جايه فقولنا هنعمل test وهنخليه خمس دقائق لو انا جيت بعد الخمس دقائق يبقى النظرية غلط ولو انا مجتش بعد الخمس دقائق فنسبنا النظرية صح
- يبقى ال p-value هي القيمه الي لو عديتها accept ال null Hypothesis
- كل اما ال p-value تقل كل اما يكون أدق

=< ال probability :

- هي بتقع تحت حاجه اسمها ال inferential statistics او الاحصاء التنبؤيه
- يعني باستخدامها عشان اعمل predictions وهي based على احصائيات عندي
- احتمال ان دينا صور يطلع من الارض مش صفر لأنه ولو كان اقرب للأستحاله لكن مفيش اي حاجه تمنع حدوثه
- في عندنا نوعين من ال probability :

Theoretical prob .1

Actual prob .2

- ال Theoretical دي عبارته عن ال prob الي بالورقه والقلم بمعنى انه مثلا احتمال كذا = احتمال كذا + احتمال كذا يعني من الآخر prob عبارته عن شوية قوانين بطبقهم وليهم اثباتات وحاجات بتثبت انها صح زي مثلا احتمالات ال coin

- ال Actual prob بتقولك انه مش معنى انك رميت coin وال prob بتاعتها 1:1 او 50:50% فده مش معناه اني لو رميتها 10 مرات فالصوره هتظهر 5 مرات والكتابه هتظهر 5 مرات لا مش شرط

- يعني مش معنى ان الحاجه احتمال حدوثها ضعيف يبقى مش هتحصل ولا لو الحاجه احتمال حدوثها كبير فده معناه انه هيجصل

- لأن كل الاحتمالات دي موجوده على الورق كقوانين انما الي بيحصل في الحياه مختلف

- يبقى انا كده مطلوب مني حاجتين اني احسب الاحتمالات واحسب نسبة حدوث كل احتمال بشكل عام

=> طيب ايه علاقة ده كله بال Ai ؟

- عندنا مبدئ مهم جدا في الحياه اسمه ال uncertainty principle مبدئ عدم التأكد
- وده بيقولنا ان كل حاجه بنحسبها وبنخططلها بتكون based على قوانين مفيش حاجه منهم اكيد
- مشكلتنا في ال Ai ان احنا بنحاول نعمل معادله بت generalize بأقل نسبة error
- يعني اساسا احنا داخليين بنحاول نعمل حاجات عارفين ان فيها احتمال خطأ وارد
- وطالما فيه نسبة error يبقى هي اقرب جدا انه فعلا تطلع الحاجه الي بنعملها دي مش صح

=>ايه هو ال Predictions ؟

- ال model اثناء ال training بيطلع المعادله الي هي المفروض اقرب solution للصح ولما باجي اعمل prediction بطلع ناتج التعويض في المعادله

دي

- الناتج الي هو اصلا برضو مش صح بس اقرب حاجه للصح
- فانا لو اديته data complicated فبالتالي انا عارفه انه هيطلع ناتج اقرب للغلط
- ودايما ال model هيطلعك الناتج وهو متأكد منه بنسبة اد ايه
- بمعنى ان انا لو اديت للمودل صورة قطة وقال لي انها قطة بنسبة 51% فده معناه ايه ؟

- ده معناه انه جاب فعلا الصح بس هو مش متأكد منه بنسبة 51 %
- فانا كده ازاي بقيس ال accuracy بتعت المودل ؟
- بأن انا بشوف كل اجابه جوابها هو متأكد منها بنسبة اد ايه
- ممكن المودل يجاوب بالصدفه البخته صح زي مثلا اني اجي اقولك اي هي عاصمة ايطاليا واديلك اربع احتمالات القاهره ولا ابو زعل ولا البدرشين ولا روما فانت سبحان الله مش عارف ان روما عاصمة ايطاليا بس عارف ان ال3 الباقيين استحاله يكونوا العاصمه بتاعتها كذلك برضو لو اديت حاجه للمودل وعمل prediction صح بنسب قليله لأنه هو متأكد انه الحاجه دي استحاله تكون من الي هو عارفهم
- لو انا اديت للمودل صورة كلب وقال لي انه قطة بنسبة 70% فده معناه ان المودل مش متعلم خالص
- ال accuracy بتاعت المودل مش يتعبر عن كل حاجه
- يعني كده ممكن يكون في مودل ال accuracy بتاعته عاليه بس هو في الحقيقه شغال غلط

=> الملخص لا تثق في ال prob هي بس طريقه لقياس ال accuracy بتاعت ال Ai وطرقه نعمل بيها predictions ونحسب نسبة نجاح ال prediction دي صح ولا غلط

=> المودل صعب انه يفضل شغال على الداتا الي هو متدرب عليها فترة كبيره لأنها الداتا في تغيير دائم

=> امتي نكتفي بال accuracy الي انا جبتها وواقف training ؟

-في عالم البيزنس 85%
-في عالم الطب والدواء 95% او في التسعينات عموما

=> ال accuracy اه مش دايما بتعبر ان المودل ماشي صح بس احنا معندناش مقياس تاني غيره

=> ممكن استخدم المودل من غير ما اعمله training اه زي ال KNN model

=> ال prob لما باجي احسب بيها ال predictions بيكون عندي 3 انواع رئيسيه

1. Dependent prob

2. Independent prob

3. Matual exexclusive

=> ال dependent prob هو احتمال حاجات انها تحصل بالاعتماد على بعضها بمعنى " ايه احتمال انه ابو تريكة يجيب جول واحتمال انه الاهلي يكون بيلعب ماهو معنى انه ابو تريكة جاب جول يبقى الاهلي بيلعب " يبقى الاتنين معتمدين على بعض وده بنشوفه في الداتا لما مثلا نقول اي احتمال انه لما ال feature الفلاني يحصله حاجه في feature تاني يتأثر ؟

=> ال Independent prob زي احتمال انه " المطره تمطر واحتمال طفل جديد يتولد في مصر " الاتنين ملهمش علاقه ببعض

=> احنا مطلوب منا اننا نشيل اي حاجه independent من ال data

=> ال mutual exexclusive هو احتمالين مينفعش يحصلوا مع بعض زي احتمال انه الاهلي يخسر واحتمال انه الاهلي يكسب في نفس الوقت

=> احنا بنستفيد بده في ايه بقى ؟

- احنا لما بنيجي في اي عمليه من عمليات ال decision making احنا بنحسب تابعات ال decision ده فمحتاج احسبها ال probability
- فانا مثلا لازم اني اتأكد ان احتمال حدوث توزيع معين هو صفر
- فال mutual exexclusive بنستخدمها لما نختار ال best decision ويكون based على الداتا الي انا طلعتها ببقى عايز حاجتين ثلاثه مثلا ميحصلوش مع بعض

- اكبر مثال لده في عملية ال drag design زي مثلا عاوزه ارفع الضغط بس مزودش السكر
- في حاجه ثاني في ال prob اسمها ال paise :
- مثلا الاهلي ايه هي كل الاحتمالات؟ انه يخسر او يكسب او يتعادل
- فال paise ده بيقولك ان كل احتمال بيكون ليه weight
- فمش لازم كل الاحتمالات تحصل بنفس النسبه

=> عندنا حاجه اسمها ال experiment ال prob بتتسب بنائ عليها فانا بعملها عشان اطلع منها كل الاحتمالات

=> عندنا حاجه اسمها ال solution space الي هي كل الاحتمالات الممكنه الي ممكن تطلع

=> عندي حاجه اسمها ال sample point ده كل احتمال لوحده يعني الاهلي يكسب ده sample point يخسر ده sample ثانيه

=> عندي event وده هو انه ال sample point تتحقق

=> يبقى احنا بنحسب ال probability بتاع event بيحصل في experiment

ال Calculus هشرحها في pdf لوحده بأمر الله دعواتكم بس بالبركة في الوقت

عن أبي أمامة الباهلي -رضي الله عنه- عن النبي -صلى الله عليه وسلم- قال: (فَضْلُ الْعَالَمِ عَلَى الْعَابِدِ كَفَضْلِي عَلَى أَدْنَاكُمْ ثُمَّ قَالَ رَسُولُ اللَّهِ -صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ-: إِنَّ اللَّهَ وَمَلَائِكَتَهُ وَأَهْلَ السَّمَاوَاتِ وَالْأَرْضِ حَتَّى النَّمْلَةُ فِي جُحْرِهَا، وَحَتَّى الْحَوْتُ لِيَصَلُّوا عَلَى مُعَلِّمِ النَّاسِ الْخَيْرِ). رواه الترمذي

