**Linear Regression**

ال linear model بيروح يجيب ال weights sum of the input features عشان نقدر نستخدمه فى إننا نعمل predict للداتا ، وعشان نعمل ال prediction محتاجين اننا نعمل train لل model عشان نقدر نحصل على ال Parameters الى بتساعد ال model يعمل generalization .

**The Normal Equation**

هى طريقة ققدر من خلالها احصل على أفضل قيم لل Parameters الى ال model من خلالها بيقدر يعمل generalization فى خطوة واحده وديه ليها مميزاتها إنها بتكون O(N) بس بالنسبه لعدد ال instances فى الداتا لكنها بتكون حوالى O(N^3) بالنسبة لعدد ال features فمثلاً لو شغال خاصة ب image classification هيكون صعب أشتغل بيها لأن عدد ال pixels of width \* height of image بيكون كبير.

**The Gradient Descent**

هنا بدل إنى اجيب القيم المثلي لل Parameters ديه فى خطوه واحده ممكن اعملها update فى خطوات متتابعه كل خطوه منها بتحاول تحسن ال Parameters ديه بحيث انها تخلى ال model يعمل generalization بشكل كويس ، فهى عبارة عن iterative process بتستمر لعدد معين من الخطوات ، والعدد ده اما يكون كبير او صغير على حسب ، وده بيتحكم فيه عامل تانى وهو ال Learning Rate ، الخطوة الى بياخدها الموديل فى كل مرة وهو بيعدل المتغيرات ديه ممكن تكون كبيرة وده ممكن يؤدى ل diverge يعنى تنتهي عند error أكبر من الى كنت بتحاول تقلله ، او انه تكون خطوات صغيرة وده بياخد وقت كتير ومحتاج تستني عدد من ال iterations كبير.

وهنا أنا بحاول اخلى ال learning Rate يكون قيمة متوسطة او مرتبطة أكتر بال task الى انا فيه من خلال انى اجرب قيم مختلفة أو انى اعمل Learning Schedule وده فيه طرق مختلفة من انى ابدء بقيمة كبيرة وبعد كده ققللها وهكذا أو العكس او انى اروح واجى بين قيمتين ما ، ويستحسن دوما إنى اخلى ال features تكون قيمها كلها فى range معين بدل ما يكون كل feature فى range مختلف عن التانى ، لان ده بيخلى ال train أسرع وأفضل .

**The type of Gradient Descent**

فى أنواع مختلفة من ال Gradient Descent وهو أنى اعمل fed لل data كلها مرة واحده فى كل iteration وده بيسمي Batch-Gradient ، أو انى اعمل fed لكل instance لوحده وده بيمسي Stochastic Gradient ، او انى اعمل fed لمجموعات من ال instances وده بيسمي Mini-Batch Gradient .

**حط الجدول الى فى ص 128**

**Polynomial features**

غالباً ما بتكون ال data فيها features كتيره ومعقدة وانى اعمل linear model مش هقدر اعمل fit بشكل كويس ، ولكن عن طريق انى اخلى فى ارتباط بين ال features ديه ومثلا انى اجيب power of two منها ال هو second degree ده هيخلى حتى لو simple model زى ال linear model ياخد باله من الارتباط ده ، ويقدر يعمل fit كويس ، وكل ما زودت عدد ال degrees ديه كل ما ده كان risk لل overfitting ، وهنا محتاج القي انسب degree تكون بتعمل generalization مع ال model .

**Over & under fitting**

بعد ما بعمل train للموديل ببتدى أشوف الموديل بتاعى ده هل فعلاً شغال كويس ولا فيه بعض المشاكل ، وده بيكون عن طريق انى اعمل validation and test قبل ما ققول والله احنا نقدر ناخد الموديل ده ونشتغل بيه فى الحياة العملية .

اول مشكلة هى انى الموديل بتاعي يكون شغال فقط كويس على الداتا الى عمل عليها train ولما اجى اعمله validation على جزء من ال data هو مشفهوش فى ال train القي نسبة error كبيرة ده بيسمي overfitting .

او انى ال model من الأول أصلاً يكون مش عارف يعمل fit لل data الى بيعمل عليها train من الأساس وده بيسمي under fitting .

وساعات ممكن يكون العيب نفسه فى ال data وده بيسمي Irreducible error . وهنا انا بحتاج اشوف انهى مشكلة من دول ال model واقع فيها ، وده ققدر اعمله من خلال ال Cross-Validation أو رسم ال Learning Curves بين ال train and validation واعمل early stopping .

فى الاخر انا بحتاج انى اظبط المشاكل ديه وده بيرجع ممكن لل model المستخدم او فى الداتا نفسها ، فمثلا بعض الحلول لل over fitting انى اعمل train على داتا اكبر واخلى الموديل يشوف instances اكتر ، او انى اضيف بعض ال regularization ، او ققلل ال degree لو كنت شغال ب polynomial model وهكذا بتختلف على حسب ال pipeline بتاعى ، وممكن يكون العيب فى الداتا نفسه وانى محتاج ابص علي ال data واصلح المشاكل الى فيها .

**Logistic Regression**

ال supervised learning فيه نوعين معروفين اما regression وهو انى اتوقع قيمة ما لكن القيمة ديه continues value ، او انها تكون قيمة من بعض القيم وده اسمه classification أو Logistic Regression ، والفرق انى بدل ما كنت بطلع النتيجة علطول زى ما فى ال Regression ، لا أنا باخد النتيجة ديه واروح اعملها map عن طريق Logistic function واخليها تكون بين 0 و 1 مثلا او على حسب وبناء علي threshold معين ، ببدء ققول لو كان ال map ده اكبر من .5 مثلا ده هيبقا تبع class معين ولو اقل هيكون تبع class اخر ، وده عباره عن Binary Classification بينما ممكن اكون بحاول اعمل classification بين أكتر من class زى مثلاً انى لما اشوف صورة رقم ققول ده انهى رقم من الارقام من 0 ل 9 مثلا وده بيسمي multi-classification .

تَمّ بِحَمدِ الله .