

۲۶ درخت تصمیم مدل ما بسیار پیچیده است و variance بالایی دارد و در حالی که bias آن
 بسیار پایین است. یعنی می توان گفت اکثر decision tree ها overfit می شوند. در تکنیک Random
 forest برای اینکه مدل ما بهتر generalized شود بهتر است که با تکنیک ها مانند bagging سعی کنیم bias-variance
 را متعادل کنیم. در تکنیک bagging ما استفاده از train sample تعدادی و با جایگزینی مبرار یک
 decision tree که نسبت به بقیه uncorrelated است در نظری می گیریم. سپس می توانیم بین نتایج های
 هر کدام از این مدل ها یک میانگین بگیریم و جواب را گزارش دهیم. باید توجه کرد چون هر درخت روی داده متفاوتی
 تصمیم می گیرد پس انتظار مدل بهتر generalized می شود. در Random forest نیز ما از تکنیک های
 مختلف bagging و boosting استفاده می کنیم حال به تاثیر پارامترهای مختلف این
 مدل می پردازیم. اولین hyperparameter تعداد درخت ها است. هر چه تعداد درخت ها بیشتر
 باشد مدل بهتر داده ها را یاد می گیرد. اما تا کردن این تعداد به دلیل computation بالا همیشه هم خوب
 نیست. متغیر جدیدی حداکثر عمق است که تعیین می کند هر کدام از درخت ها تا
 چه عمقی ادامه می دهند. متغیر دیگر حداقل داده در یک مغلی است. در واقع ما نگاه می کنیم که اگر به آن تعداد
 sample رسیدیم دیگر متوقف شویم و مقایسه با درخت تصمیم را ادامه ندهیم. متغیر آخر هم
 حداکثر ویژگی ها است. هر بار به هر درخت تعداد محدودی ویژگی می دهیم تا به صورت uncorrelated
 تشکیل شود. با استفاده از tune کردن این hyperparameter ها می توان یک RF بسیار خوب
 ساخت که trade-off بین variance و bias را رعایت می کند. به عبارتی از decision tree های
 overfit شده می توان با تکنیک های مختلف که بالا ذکر شد به مدل هایی رسید که بسیار خوب و دانه ها
 generalize می کنند.