

- 1: تمامی کتابخانه های مورد نیاز را فراخوانی میکنیم.
- 2: فراخوانی داده ها از فایل دیتاها و جایگزینی مقادیر گم شده در فیچرها با مقادیر میانگین سایر مشاهدات هر فیچر سپس حذف مشاهداتی که لیبل نامعلوم دارند (تعداد 566 مشاهده از 7831 مشاهده). انتخاب تصادفی 80 درصد دیتاها به عنوان داده آموزشی، 10 درصد به عنوان داده آزمایشی و 10 درصد داده اعتبار سنج.
- 3: استفاده از روش Grid Search CV جهت پیدا کردن بهترین هایپرپارامترها در شبکه با یک لایه پنهان .
مرحله اول: ساختن مدل در تابعی تحت عنوان `cteate_model_for_GS` با مقادیر اولیه دلخواه هایپرپارامترها.
مرحله دوم: استفاده از کتابخانه Keras و مشخص کردن Optimizer مدل و تعداد neurons لایه پنهان با استفاده از مدل تعریف شده در قسمت قبل و 5k-cross-validation. بدست آوردن دقت این مدل بر روی دیتاهای آزمایشی
استفاده از Optimizer و تعداد neurons لایه پنهان بدست آمده در قسمت قبل به جای مقادیر اولیه دلخواه در مدل سپس مشخص کردن `init_mode` لایه نهان و لایه خروجی و `dropout_rate` پس از لایه پنهان با استفاده از مدل جدید و 5k-cross-validation. بدست آوردن دقت این مدل بر روی دیتاهای آزمایشی.
- مرحله سوم: ساختن مدل نهایی با توجه به پارامترهای بدست آمده مرحله قبل، بررسی دقت و خطای مدل روی دیتاهای آموزشی و اعتبارسنج، نمودارهای مدل و دقت و خطای مدل بر روی داده های آزمایشی – یافتن بهترین epoch با استفاده از نمودارها و اجرای مدل با این مقدار مناسب و بدست آوردن دقت و خطای مدل روی دیتاهای آموزشی و اعتبارسنج، نمودارهای مدل و دقت و خطای مدل بر روی داده های آزمایشی
- 4 : استفاده از روش Keras Tuning جهت پیدا کردن بهترین پارامترها در شبکه با یک و دو لایه پنهان .
برای هر دو مدل شبکه عصبی با یک و دو لایه پنهان بترتیب مراحل زیر طی میشود:
مرحله اول: ساختن مدل بر روی داده های آموزشی در تابعی تحت عنوان `cteate_model_for_KT` .
مرحله دوم: استفاده از کتابخانه Keras_tuner و بدست آوردن بهترین هایپرپارامترها با استفاده از داده های اعتبارسنج - بررسی دقت مدل بر روی داده های آزمایشی
مرحله سوم: ساختن مدل نهایی با توجه به پارامترهای بدست آمده مرحله قبل، بررسی دقت و خطای مدل روی دیتاهای آموزشی و اعتبارسنج، نمودارهای مدل و دقت و خطای مدل بر روی داده های آزمایشی – یافتن بهترین epoch با استفاده از نمودارها و اجرای مدل با این مقدار مناسب و بدست آوردن دقت و خطای مدل روی دیتاهای آموزشی و اعتبارسنج، نمودارهای مدل و دقت و خطای مدل بر روی داده های آزمایشی.
- 5: اجرای bagging بر روی بهترین مدل بدست آمده از قسمت 3 و 4
(به دلیل طولانی بودن اجرا و پیچیدگی بیش از حد استفاده از epoch=350 و تفاوت ناچیز در دقت و خطا. از epoch=50 استفاده شده.)
محاسبه دقت و خطای مدل با همه داده ها جز داده های آزمایشی -محاسبه دقت و خطای مدل بر روی داده های آزمایشی
اجرای bagging با جداکردن داده های out of bag پس از مدل سازی و محاسبه دقت بر روی این داده ها-محاسبه دقت مدل بر روی داده های آزمایشی
- 6: انتخاب 10 داده ابتدایی از داده های آزمایشی و مشاهده لیبل اصلی و لیبل های پیش بینی شده آنها توسط مدلهای زده شده و مقایسه مدلهای.

در این فایل از سه روش انتخاب متغیر استفاده شده و سپس روش Keras Tuning طبق قبل و مراحل گفته شده انجام میشود. انتخاب 10 داده ابتدایی از داده های آزمایشی متناسب با هر نوع انتخاب متغیر و مشاهده لیبیل اصلی و لیبیل های پیش بینی شده آنها توسط مدل های زده شده و مقایسه مدل ها.

جدول مدل ها:

Model	epoch	Number hidden layer	loss	accuracy	val_loss	val_accuracy	test set loss	test set accuracy
GS_modelbest	350	1	0.0889	0.9748	0.0895	0.9786	0.12978	0.9683
KT_model_1best	32	1	0.0923	0.9748	0.0887	0.9786	0.1261	0.9683
KT_model_2best	25	2	0.0878	0.9742	0.0958	0.9817	0.1345	0.9669
KT_model_f1	500	1	0.1049	0.9733	0.0857	0.9801	0.1210	0.9669
KT_model_ch1best	30	1	0.0909	0.9743	0.0923	0.9832	0.1287	0.9669
KT_model_mi1best	420	1	0.1231	0.9708	0.0899	9801	0.1247	0.9683
KT_model_f2best	70	2	0.0949	0.9737	0.0889	0.9786	0.1220	0.9683
KT_model_ch2best	25	2	0.0944	0.9730	0.0911	0.9801	0.1287	0.9669
KT_model_mi2best	25	2	0.0890	0.9754	0.0903	0.9832	0.1311	0.9669

جدول ensemble ها:

model	epoch	Number hidden layer	test set accurac y	accurac y for out-of- bag
bag_clf_GS	50	1	0.9573	-
outbag_clf_GS	50	1	0.9573	0.9574
bag_clf_KT1	32	1	0.9642	-
outbag_clf_KT1	32	1	0.9656	0.9732
bag_clf_KT2	25	2	0.9656	-
outbag_clf_KT2	25	2	0.9656	0.9733