

فایل final_NNSider

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 لیبیل نامعلوم دارند (تعداد 566 مشاهده از 7831 مشاهده).

3: استفاده از روش Grid Search CV جهت پیدا کردن بهترین هایپر پارامترها در شبکه با یک لایه پنهان .

انتخاب تصادفی 90 درصد دیتاها به عنوان داده آموزشی، 5 درصد به عنوان داده آزمایشی و 5 درصد داده اعتبار سنج (این نوع انتخاب سایز مجموعه داده ها با توجه به تجربیاتی که بدست آمد انجام شد و انتخاب بیشتر داده ها به عنوان تست و اعتبار سنج مفید نبود)

مرحله اول: ساختن مدل در تابعی تحت عنوان `create_model_for_GS` با مقادیر اولیه دلخواه هایپر پارامترها.

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 دیتاها ی آزمایشی

استفاده از Optimizer و تعداد neurons لایه پنهان بدست آمده در قسمت قبل به جای مقادیر اولیه دلخواه در مدل سپس بدست مشخص کردن `init_mode` لایه نهان و لایه خروجی و `dropout_rate` پس از لایه پنهان لایه 1 استفاده از مدل جدید و `5k-cross-validation`. بدست آوردن دقت این مدل بر روی دیتاها ی آزمایشی.

مرحله سوم: ساخت مدل نهایی با توجه به پارامترهای بدست آمده مرحله قبل، بررسی دقت و خطای مدل روی دیتاهای آموزشی و اعتبار سنج، 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

انتخاب تصادفی 90 درصد دیتاها به عنوان داده آموزشی، 10 درصد داده اعتبار سنج (این نوع انتخاب سایز مجموعه داده ها با توجه به تجربیاتی که بدست آمد انجام شد و دلیل اینکه داده تست در نظر نگرفتیم این بود که تجربه نشان داد دقت اعتبار سنج کم میشود وقتی بخشی را برای آزمایش در نظر بگیریم) .

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 `create_model_for_KT` .

مرحله دوم: استفاده از کتابخانه Keras_tuner و بدست آوردن بهترین هایپر پارامترها با استفاده از داده های اعتبار سنج

مرحله سوم: ساخت مدل نهایی با توجه به پارامترهای بدست آمده مرحله قبل، بررسی دقت و خطای مدل روی دیتاهای آموزشی و اعتبار سنج، نمودارهای مدل.

5: اجرای bagging بر روی بهترین مدل بدست آمده از قسمت 3 و 4

GS : 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 جدا کردن 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

KT : محاسبه دقت و خطای مدل -محاسبه دقت و خطای مدل بر روی داده های اعتبار سنج که از آنها استفاده ای در bagging نشد و میتوان به آنها به چشم داده آزمایشی نگاه کرد.

اجرای bagging با جدا کردن داده های out of bag پس از مدل سازی و محاسبه دقت و خطای مدل بر روی داده های اعتبار سنج که از آنها استفاده ای در bagging نشد و میتوان به آنها به چشم داده آزمایشی نگاه کرد.

6:

انتخاب 10 داده ابتدایی از داده های آزمایشی و مشاهده لیبل اصلی و لیبل های پیش بینی شده آنها توسط مدل های زده شده با GS و مقایسه مدل ها.

انتخاب 10 داده ابتدایی از داده های اعتبار سنج و مشاهده لیبل اصلی و لیبل های پیش بینی شده آنها توسط مدل های زده شده با bagging KT و مقایسه مدل ها.

برای KT چون داده آزمایشی در نظر نگرفتیم قسمت پیش بینی نداریم.

فایل final_sele_NNSider

در این فایل از سه روش انتخاب متغیر استفاده شده و و سپس روش Keras Tuning طبق قبل و مراحل گفته شده انجام میشود.

انتخاب 10 داده ابتدایی از داده های اعتبار سنج متناسب (به عنوان دیتای تست به دلیل گفته شده) با روش انتخاب متغیر Ch2 و مشاهده لیبل اصلی و لیبل های پیش بینی شده آنها توسط bagging های زده شده و مقایسه آنها.

جدول مدل ها:

		Number hidden layer	loss	accuracy	val_loss	val_accuracy	test set loss	test set accuracy
GS_model	50	1	0.1854	0.9456	0.7668	0.6765	0.7461	0.7361
KT_model_1	50	1	0.5604	0.7196	0.6028	0.7203	-	-
KT_model_2	50	2	0.5516	0.7212	0.6084	0.6853	-	-
KT_model_f1	50	1	0.5837	0.6963	0.6016	0.7063	-	-
KT_model_ch1	50	1	0.5780	0.7033	0.6119	0.6573	-	-
KT_model_mi1	100	1	0.5479	0.7243	0.6114	0.6783	-	-
KT_model_f2	50	2	0.5760	0.6955	0.5861	0.7203	-	-
KT_model_ch2best	25	2	0.0944	0.9730	0.0911	0.9801	-	-
KT_model_mi2	50	2	0.5751	0.7025	.5997	0.6713	-	-

جدول ensemble ها:

model	epoch	Number hidden layer	test set accuracy	accuracy for out-of-bag	val set accuracy
bag_clf_GS	50	1	0.70833	-	-
outbag_clf_GS	50	1	0.70833	0.6878	-
bag_clf_KT1	50	1	-	-	0.6713
outbag_clf_KT1	50	1	-	0.63785	0.6713
bag_clf_KT2	50	2	-	-	0.6853
outbag_clf_KT2	50	2	-	0.6573	0.6783
bag_clf_KTch1	50	1	-	-	0.6853
outbag_clf_KTch1	50	1	-	0.6495	0.6853
bag_clf_KTch2	50	2	-	-	0.67832
outbag_clf_KTch2	50	2	-	0.6495	0.6993

همانگونه که از تحلیل دقت مدلها پیداست بهترین مدل **KT_model_ch2best** است که بالاترین دقت روی آموزش و دیتاهای اعتبارسنج را دارد و شکل نمودار learning curve ان بصورت زیر است.

