

RoBERTa training (1290 examples): cross-entropy loss

