1512 Project

(a) What is the purpose of using KD in this paper? [0.5 Marks]

(b) In the paper, what knowledge is transferred from the teacher model to the student model?

[0.5 Marks]

(c) What is the temperature hyperparameter *T*? Why do we use it when transferring knowledge

from one model to another? What effect does the temperature hyperparameter have in KD?

[0.5 Marks]

(d) Explain in detail the loss functions on which the teacher and student model are trained in

this paper. How does the task balance parameter affect student learning? [0.5 Marks]

(e) Can we look at the KD as a regularization technique, here? Explain your rationale. [0.5

Marks]