

## 1. Why RNN is useful?

الثروة ثابتة في المدى، ولكنها متغيرة في المدى القصير.

RNN eines sehr großen (ca. 200000 MLP-neur.) Platzes  
mit großer Variabilität und Fluktuationen.

RNN performs well on time-series data with multi-step predictions.

RNN. in the middle of the block of apartments was a bank  
which shared its walls with the other buildings.

دروایز شکری و شکری خوشبختانه با CNN میتوانند  
که نتیجه این ساختار بگوییم

کیونا در حین مطریزمانی برای تمثیل

2. How to prevent Vanishing gradient problem?

3. How seq2seq models are useful?

پریمیو سکوئینس تو سکوئینس  
Sequence-to-Sequence  
وایلر خروجی سکوئینس  
وایلر خروجی سکوئینس

4. Why do we need gated recurrent unit?

حافظه‌گذاری LSTM و GRU با استفاده از متریک داده همچوں بزرگواریون یا انتشاریه

*Fig. 1. A photograph of the same three species of *Leucosoma* as shown in Figure 1, but taken from a different angle.*

دسته هایی از حالتی LSTM را که در  
نهاشی از مدل GRU و LSTM

5. Which of the following are true about Batch Normalization?

6. Why is scaling ( $\gamma$ ) and shifting ( $\beta$ ) often applied after the standard normalization in the batch normalization layer?

Normalisierung mitgliedert die Daten in gleichmäßige Teile, die leichter verarbeitet werden können.

F. Please describe F1 score and AUC?

تَذَكِّرْ فِي recall، وَتَعْلَمْ فِي learning objects. F1-score

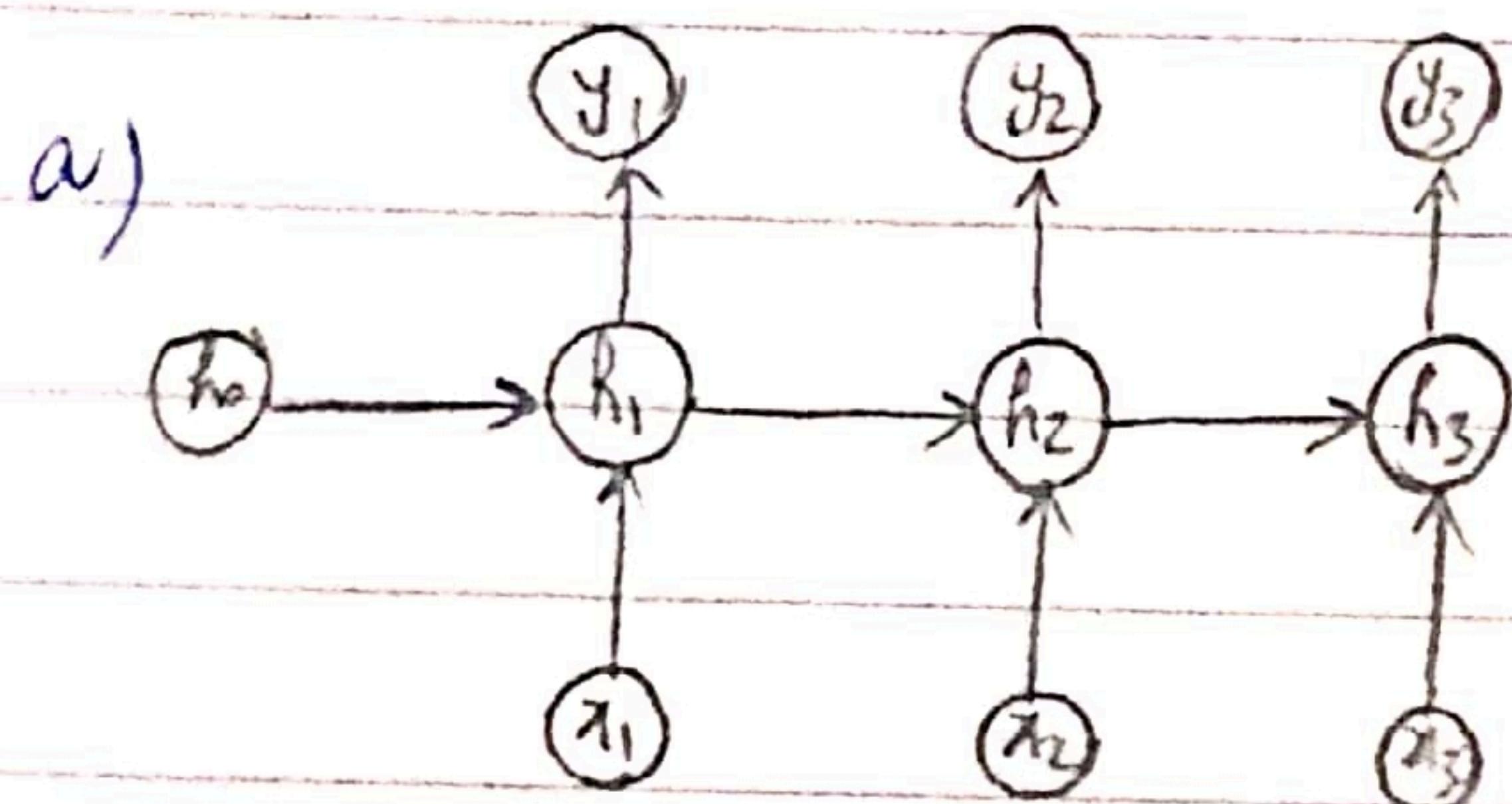
$$F_1\text{-score} = \frac{2 \times (\text{Precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$

کانول زنگنه ای مارکسیست روسیه تواند AUC  
خود را برای پیشگیری از بدبختی خود را  
با خود در همان حالت داشت. این مقاله با  
AUC را در این ایام ایجاد کرد. این مقاله  
با نظرخواهی از ایوان ایوانوف (وزیر امور

8. Which of the following can be significant causes of the vanishing gradient problem in Deep Learning?

a, c, d

9.



b)  $\text{Loss} = \frac{1}{2} (s-y)^2 \xrightarrow{s=a^h_T} \text{Loss} = \frac{1}{2} (ah_T - y)^2$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial w} = [a \cdot (ah_T - y)] \cdot [h_{+1} \cdot (1 - \tanh^2(wh_{+1} + cw_{+1}))]$$

$$\left( \frac{\partial h}{\partial w} = 2 \cdot \frac{1}{2} \cdot a(ah_T - y), \frac{\partial h}{\partial w} = (1 - \tanh^2(wh_{+1} + cw_{+1})) \cdot h_{+1} \right)$$

$$w = w_0 + \eta \cdot \frac{\partial L}{\partial w}$$

c) The answer is the same as the previous part.

d) *الإجابة تختلف باختلاف المعلمات مثل نسبت المعلمات، حيث إن المعلمات التي تؤدي إلى انخفاض التباين تؤدي إلى تباطأ التعلم، مما يزيد من وقت التعلم.*