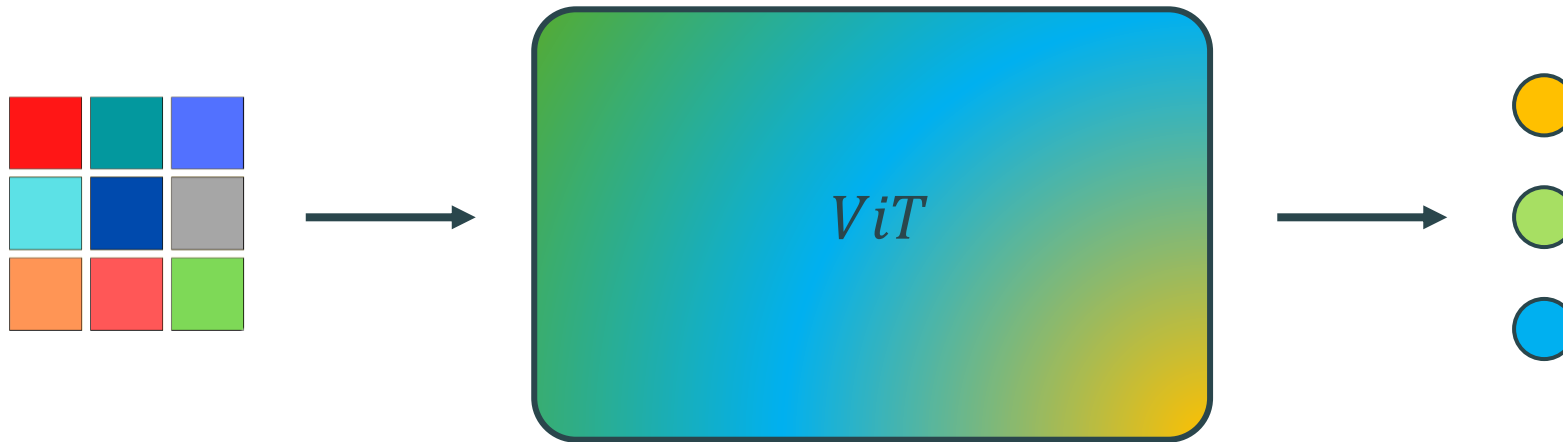


دوره در قلب Transformers (Dive Into Transformer)

جلسه اول : Vision Transformer



حامد قاسمی

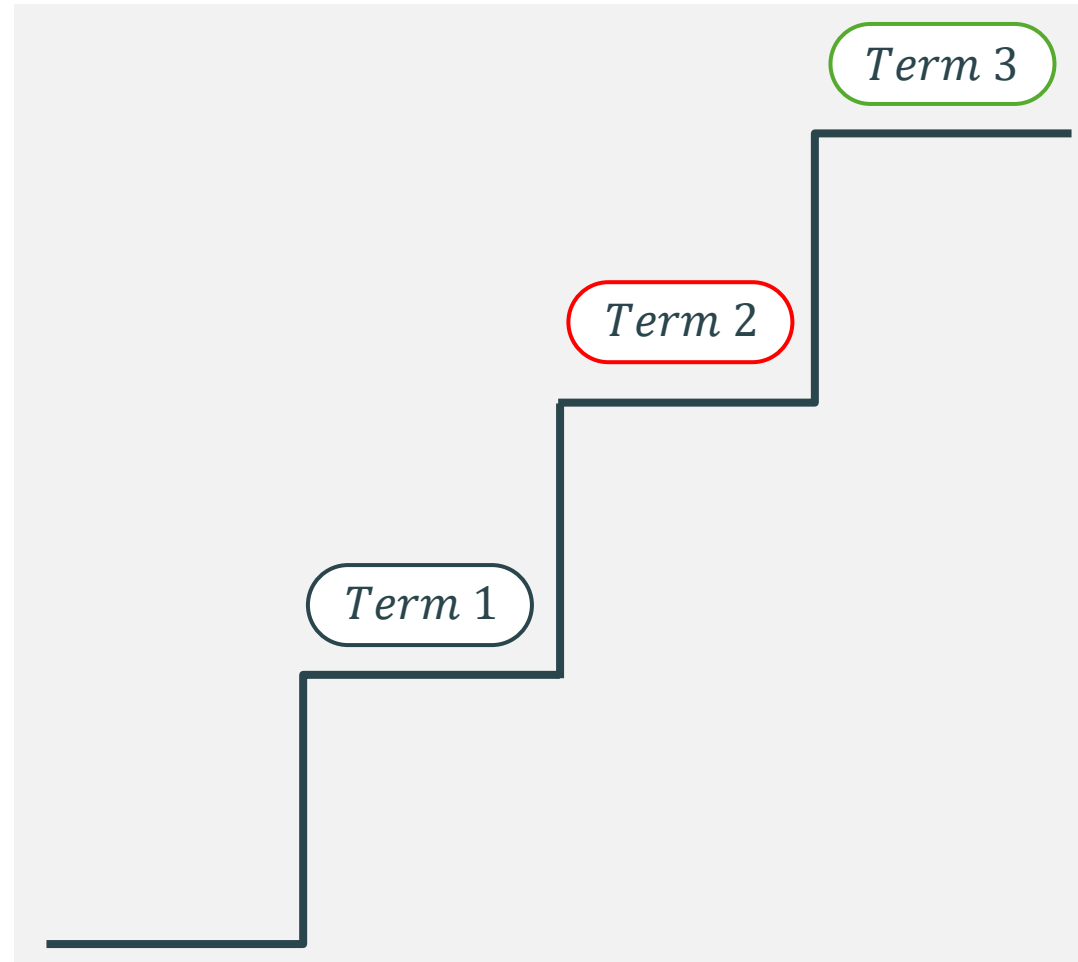
دانشجوی دکتری هوش مصنوعی دانشگاه تهران



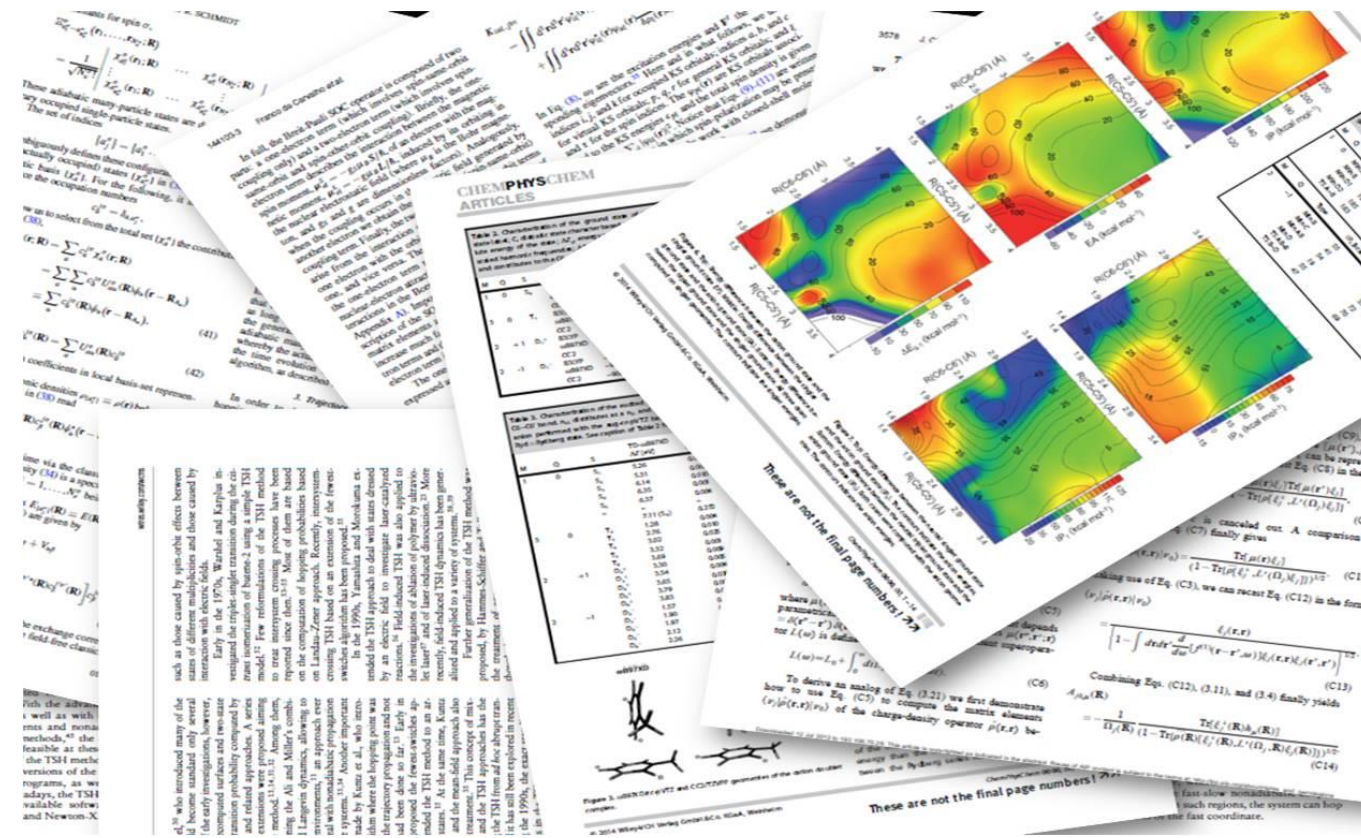
اما کمی حرف ...

دوره سخته ؟

کجاییم؟



اتفاق مهم ۱: از اسلاید خبری نیست. مستقیم به سراغ مقالات خواهیم رفت.



نتیجه مهم : **حتما دو بار ویدیو** هر جلسه را ببینید.

اتفاق مهم ۲ : یک خط از کدهای رسمی Keras عوض نمی شوند.

```
ch
ch import nn
ch import optim
orchvision.transforms import transforms
t numpy as np
PIL import Image
ort matplotlib.pyplot as plt

ebm tqdm.notebook import tqdm
ngport warnings
lternings.filterwarnings("ignore")

m = transform = transforms.Compose([
notebook
arnings
.filterwarni.img = 3072

import transforms.Resize((32, 32))
```

the two distributions, i.e. $D(\mathbf{x}) = \frac{1}{2}$.

Algorithm 1 Minibatch stochastic gradient descent training of generative steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$, experiments.

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise distribution $p_{\text{noise}}(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log (1 - D(G(z^{(i)})))]$$

of 11

<

>

Q

1502...

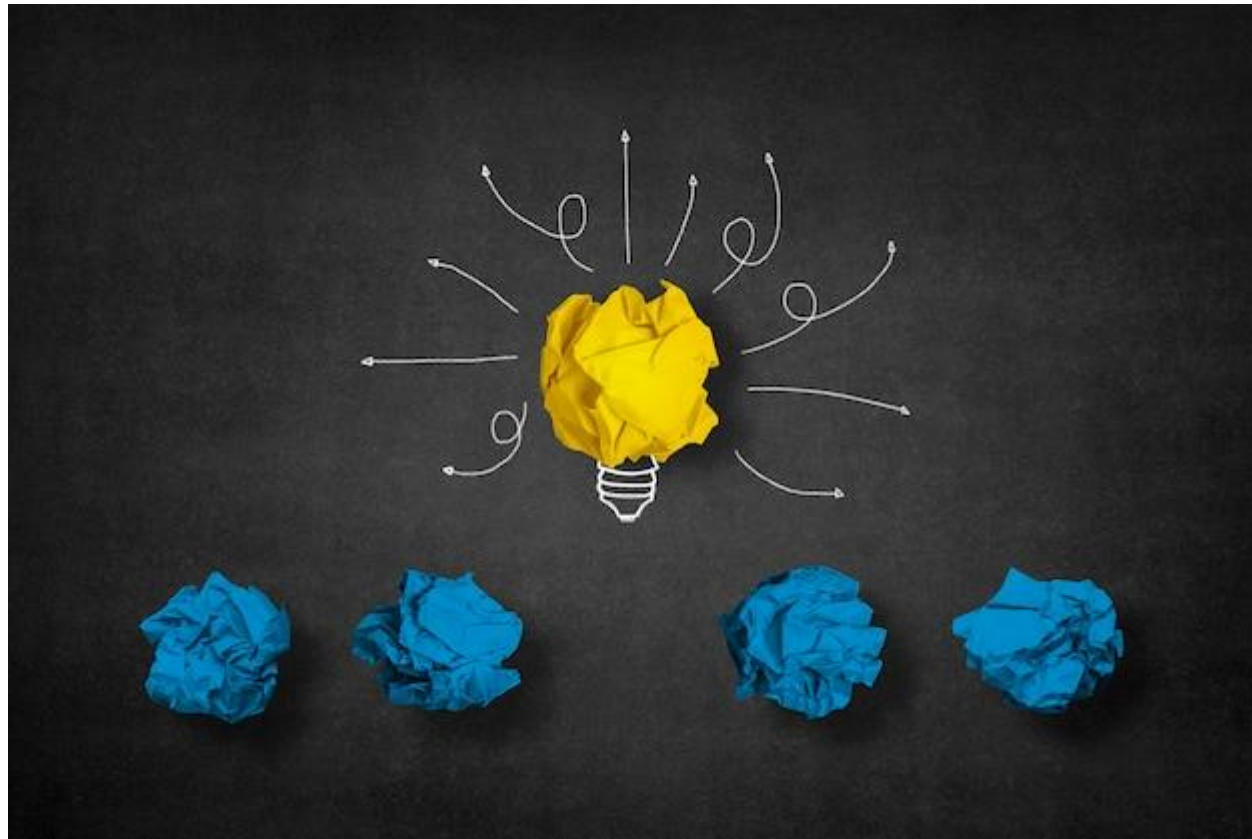
98.27%

Q

Q

vector, and \mathcal{X} be the set of these inputs over the training data set. The normalization can then be written as a trace over the data set. we introduce, for each $x^{(i)}$ we introduce

اتفاق مهم ۳ : تحلیلی و نقادانه به مقالات و کدها نگاه کنیم.



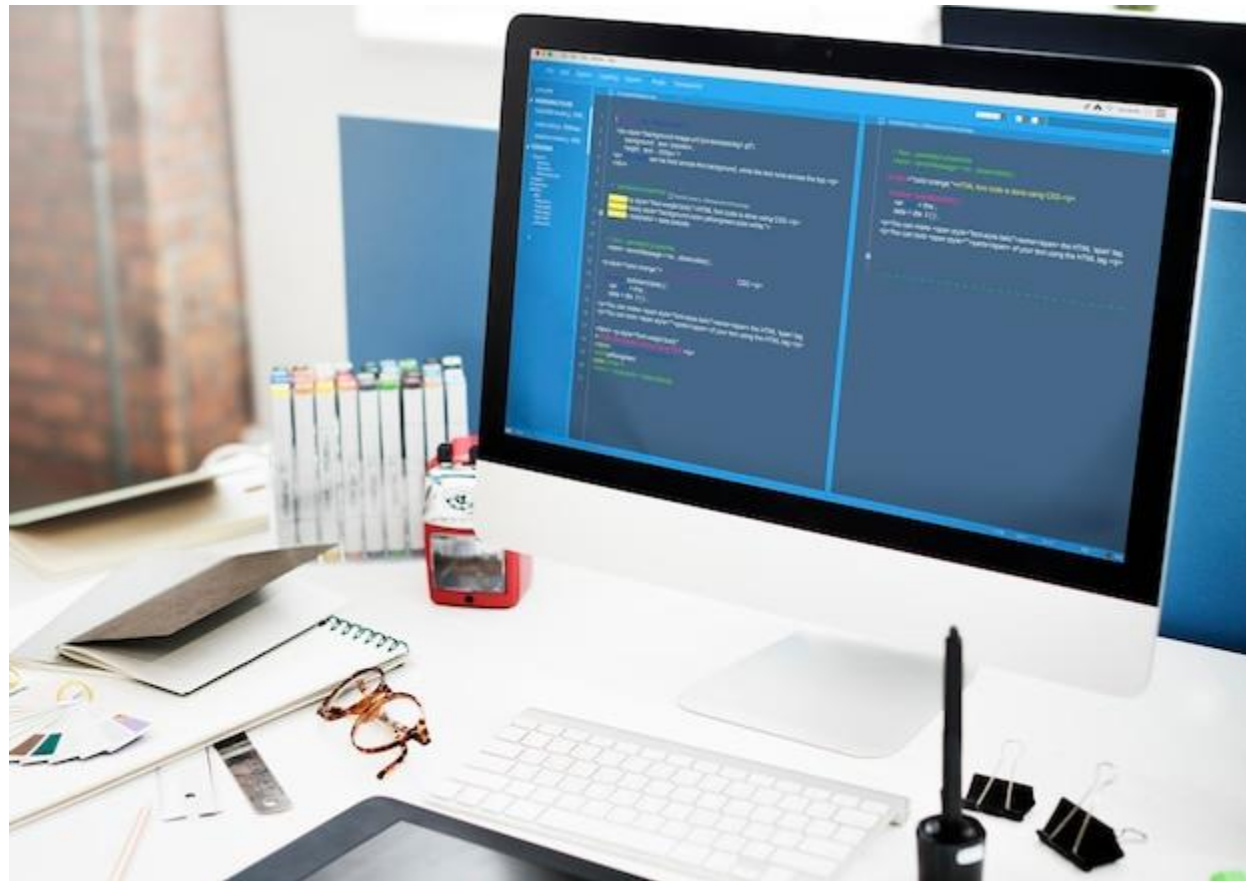
اتفاق مهم ۴: تمرینات هم متفاوت می شود.

We also evaluate on the 19-task VTAB classification suite (Zhai et al., 2019b). VTAB evaluates low-data transfer to diverse tasks, using 1 000 training examples per task. The tasks are divided into three groups: *Natural* – tasks like the above, Pets, CIFAR, etc. *Specialized* – medical and satellite imagery, and *Structured* – tasks that require geometric understanding like localization.

Model Variants. We base ViT configurations on those used for BERT (Devlin et al., 2019), as summarized in Table 1. The “Base” and “Large” models are directly adopted from BERT and we add the larger “Huge” model. In what follows we use brief notation to indicate the model size and the input patch size: for instance, ViT-L/16 means the “Large” variant with 16×16 input patch size. Note that the Transformer’s sequence length is inversely proportional to the square of the patch size, thus models with smaller patch size are computationally more expensive.

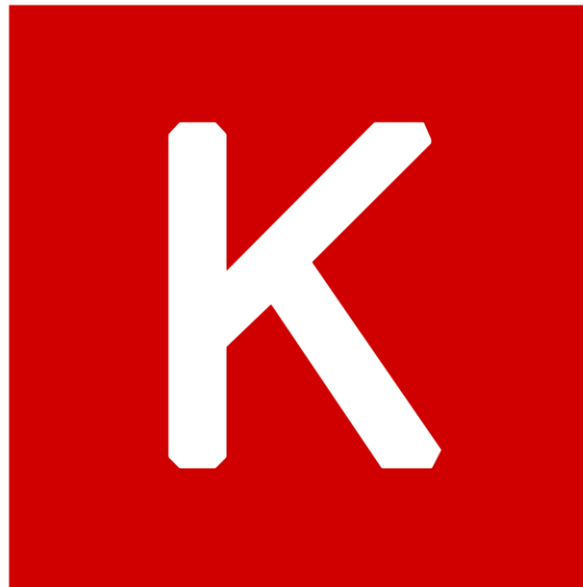
اگر سوالی دارید بفرمایید و اگر نه بریم **سر وقت مقاله**

و اما برویم سر وقت **کد**

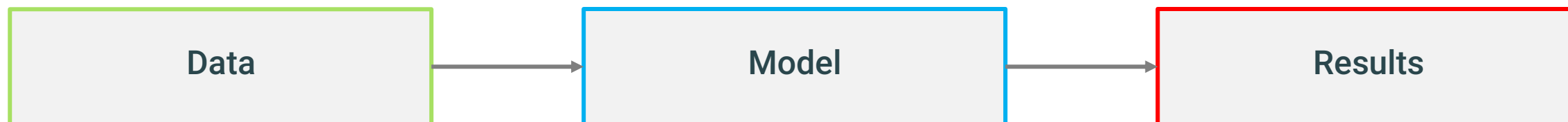


قبلش چند نکته رو مرور کنیم.

نکته اول : تمامی کدهایی که در این دوره بررسی می شوند زیر ۳۰۰ خط هستند !



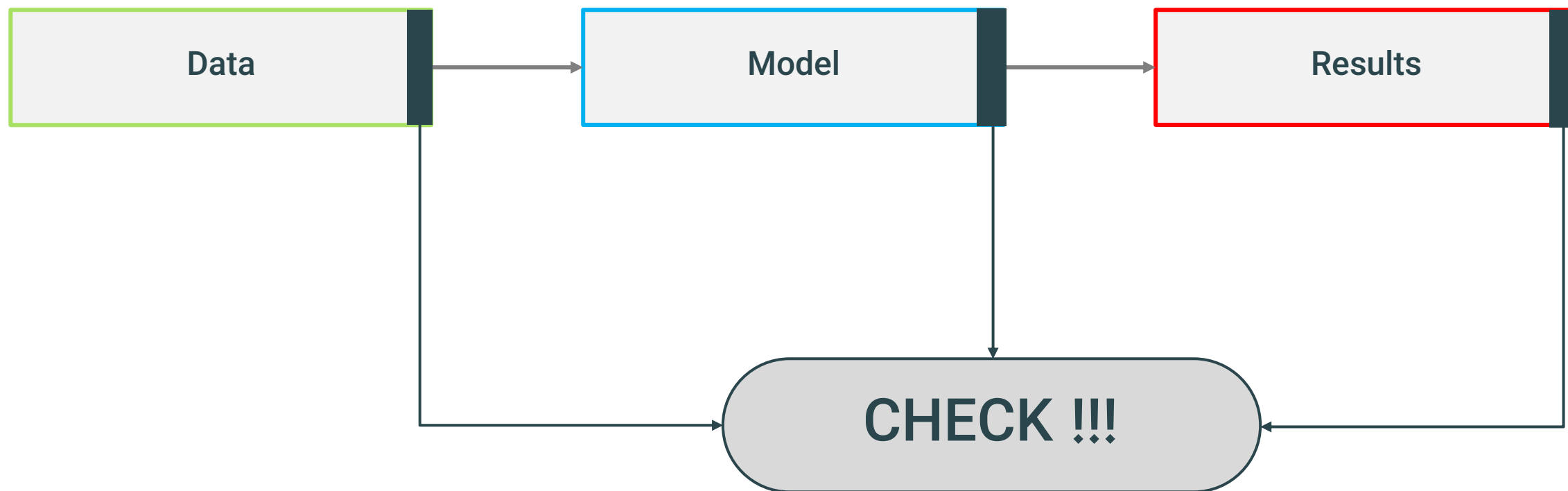
بچه ها برید بالا ، بیایید پایین، **کد شما سه بخش داره :**



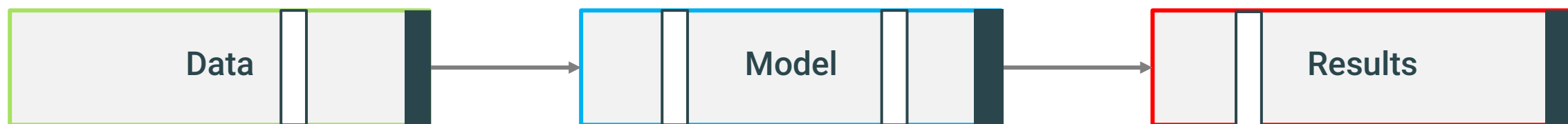
نکته مهم ۱ : سعی کنید هنگام نوشتن کد این ترتیب را رعایت کنید.



نکته مهم ۲ : بعد از هر مرحله ، درست بودن آن مرحله را **check** کنید.



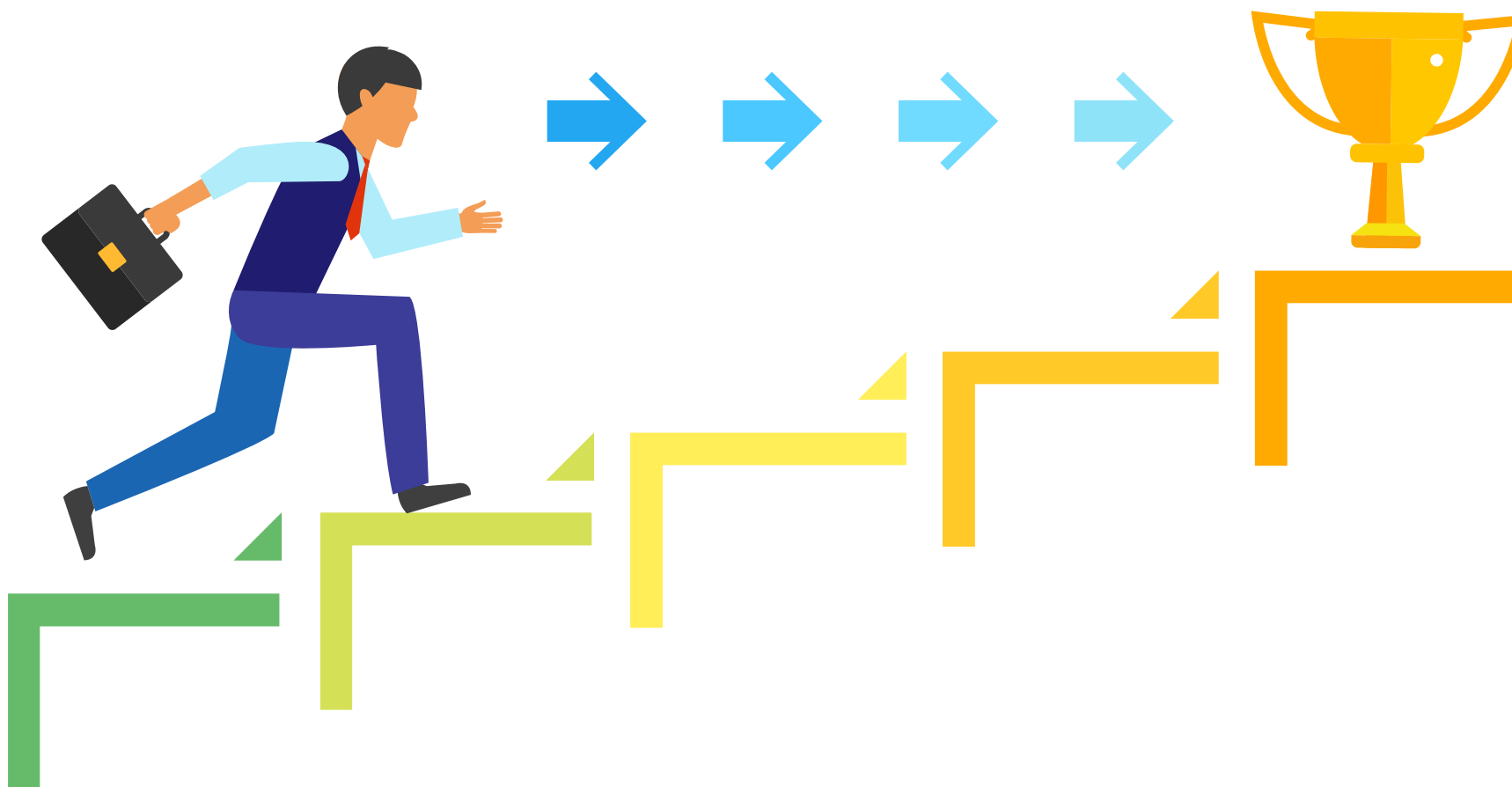
نکته مهم ۳ : ممکنه در مراحل میانی نیز شما **check** نیاز داشته باشید.



کل حرف این است : هر آنچیزی که **مینویسد** را **check** کنید در هر مرحله



نکته مهم بعدی : از کوچکترین بخشی که میتوانید آغاز کنید. به کل قصه نگاه نکنید !



قرار نیست ابتدای راه ، تا انتها همه چیز مشخص باشد ! فانوسی به دست بگیرید و در این تاریکی حرکت کنید !



پیاده سازی مقالات فرآیندی است که با **تمرین ، تکرار و تلاش** بدست می آید.



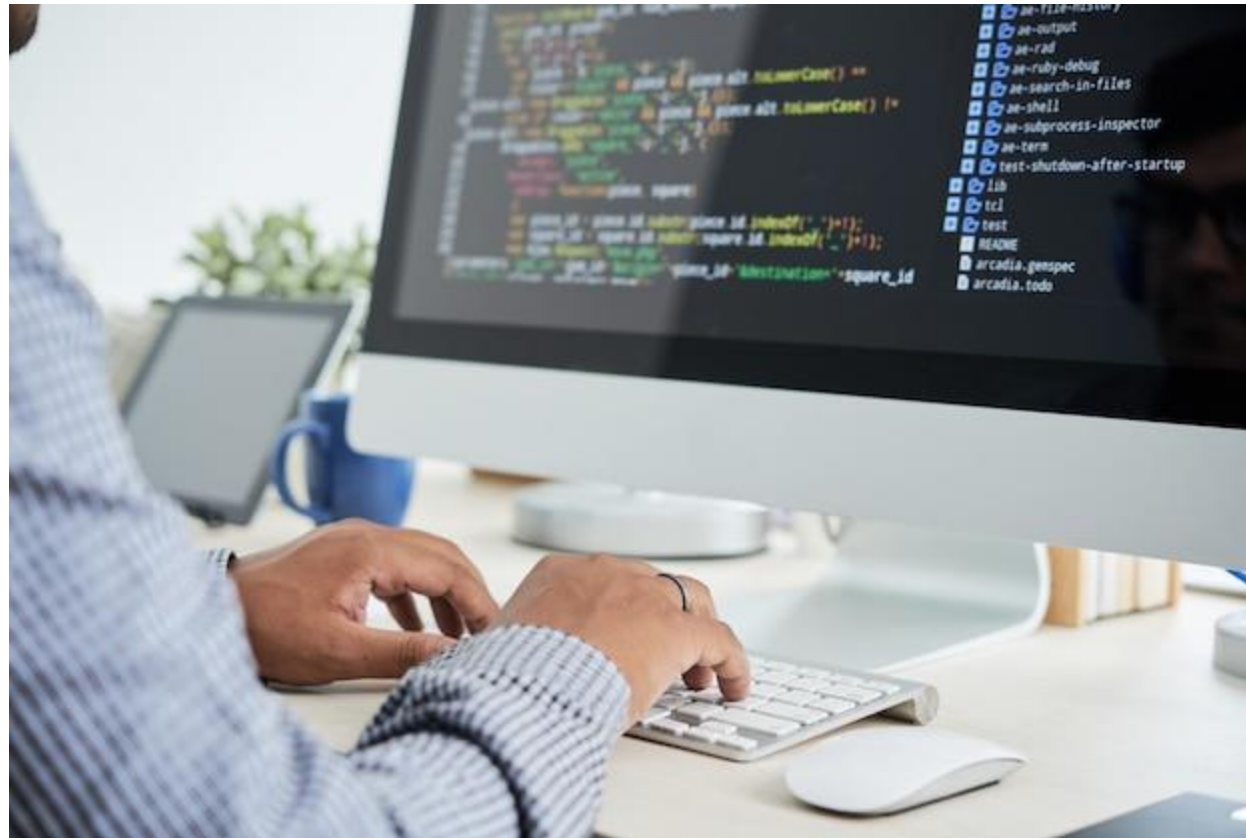
هر چه جلوتر بروید ، سنگلاخ های مسیر برای شما کم و کمتر می شود و اعتماد به نفس شما هم افزایش می یابد.



سعی کنید از کدهای مختلف **تا میتوانید یاد بگیرید. تا میتوانید !**



کدها نوشته می شوند. وحی نمی شوند.



سعی من این هست که به کدها **نقادانه** نگاه کنیم. **چرایی** وجود هر بخش و هر خط **کد** را درک کنیم.



اما برویم سر وقت کد Vision Transformer



خیلی با **صبر و حوصله** هم کد بنویسید و هم کدها رو بررسی کنید.



فقط نخوايد تموم بشه. فقط نخوايد ران كنيد. فقط نخوايد نتيجه بگيريد.

عميق بشيد. خطاها رو بفهميد و در راستاي رفع خطاها تلاش كنيد.

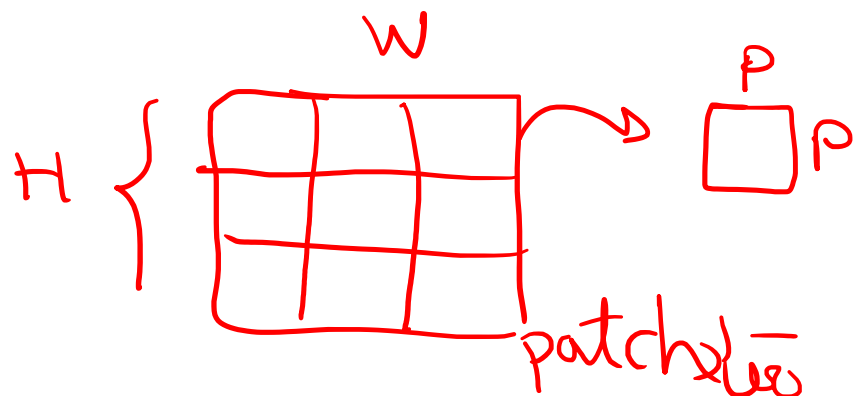


وقتی صبر و حوصله دارید و یاد میگیرید از کدها اون موقع هست که کدها کم کم دیگه سخت نیستند هر بار ...



پیچوندن یا همه جزییات را فهمیدن ؟ تعادل جایی در این بین است !





$$= \frac{HW}{p^2}$$

$$= \frac{H \times H}{p^2}$$

$$= \left(\frac{H}{p}\right)^2$$

Softmax \rightarrow $\frac{e^{-L_i}}{\sum_j e^{-L_j}}$

اگر $L_i \rightarrow -\infty$ $\rightarrow \infty$

$$\frac{x - \mu}{\sqrt{\sigma^2}}$$

$$\begin{aligned} -\infty & \rightarrow 0 \\ -\infty & \rightarrow 0 \\ +\infty & \rightarrow 0 \\ 1 & \rightarrow 0 \end{aligned}$$