**Problem Setup**

What is NAS (network architecture search)? It aims to automatically select a proper operation from an operation set for each edge in a dense graph.

DARTS converts discrete operation selection into continuously weighting a set of operations.

**Main Results**

**Gradient descent for optimization:**

inner optimization: \( W^1_t(k+1) = W^1_t(k) - \alpha \nabla \mathcal{L}(W_t, \alpha) \)

outer optimization: \( \alpha^t(k+1) = \alpha^t(k) - \alpha \nabla \mathcal{L}(W^1_t, \alpha) \)

**Convergence of inner optimization.** Under proper assumptions, for inner problem, the gradient descent algorithm can enjoy linear convergence rate:

\[ F_{\alpha}(W(k+1), \alpha) \leq (1 - \lambda) F_{\alpha}(W(k), \alpha) \]

where \( \lambda = c_0 \sum_{t=1}^{\infty} \left( \alpha^t \right)^2 \), in which \( \alpha^t \) and \( \alpha \) respectively denote weights of skip and convolutional connections, is a constant and \( \alpha \) is learning rate.

**Convergence Comparison.** With proper assumptions, shallow network \( B \) can converge faster than the deep network \( A \).

**Observations:** dominated skip connections in the architectures selected by DARTS

**Problem:**

1) why DARTS prefers to select so many skip connections?

2) how to avoid the dominated skip connections?

**Solution to Alleviate Unfair Competition among Operations**

**Step 3.** divide operations in network into skip connection and non-skip connection groups, compute their average gate activation probabilities:

\[ \mathcal{L}_{\text{skip}}(\beta) = \sum_{\mathcal{S}} \mathbb{E} \left[ \beta_{\mathcal{S}} \right] - \ln \left( \frac{2}{\beta_{\mathcal{S}}} \right) \]

**Step 4.** penalize \( \mathcal{L}_{\text{skip}}(\beta) \) and \( \mathcal{L}_{\text{non-skip}}(\beta) \) independently to avoid competition between skip connection and other type operations.

**Issue 2 of independent gate:** searching algorithms prefer to select shallow networks due to their faster convergence rate over deep ones.

**Final Model:**

\[ \min_{\beta, W} \mathcal{L}_{\text{skip}}(\beta) + \lambda \mathcal{L}_{\text{non-skip}}(\beta) \]

**Experiments**

- CIFAR10: 99.3% test accuracy on efficient model
- CIFAR100: 74.3% test accuracy on efficient model
- ImageNet: 73.6% test accuracy on efficient model