/Users/wenqian/semantic-cache/venv/bin/python /Users/wenqian/semantic-cache/paras\_evaluate.py

==================================================

Running dimension × threshold comparison...

==================================================

Testing 256d - Strict (θ1=0.95, θ2=0.9, mode=global)

θ1=0.95, θ2=0.9, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.980)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: llm (0.000)

[session\_000] Q5: llm (0.000)

[session\_001] Q1: cache (0.967)

[session\_001] Q2: llm (0.000)

[session\_002] Q1: llm (0.000)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: llm (0.000)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.982)

[session\_003] Q3: llm (0.000)

[session\_003] Q4: cache (0.976)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.977)

[session\_004] Q2: llm (0.000)

[session\_005] Q1: llm (0.000)

[session\_005] Q2: cache (0.992)

[session\_005] Q3: cache (0.982)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: cache (0.959)

[session\_006] Q2: llm (0.000)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: cache (0.979)

[session\_007] Q3: cache (0.975)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: llm (0.000)

[session\_008] Q1: llm (0.000)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: llm (0.000)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.987)

[session\_009] Q3: cache (0.980)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: llm (0.000)

[session\_010] Q3: llm (0.000)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.983)

[session\_011] Q3: cache (0.975)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: llm (0.000)

[session\_012] Q3: llm (0.000)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: llm (0.000)

[session\_013] Q1: cache (0.989)

[session\_013] Q2: cache (0.980)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: llm (0.000)

[session\_014] Q3: cache (0.984)

[session\_014] Q4: llm (0.000)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: llm (0.000)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: llm (0.000)

[session\_016] Q3: llm (0.000)

[session\_016] Q4: llm (0.000)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: cache (0.950)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 256d - Balanced (θ1=0.9, θ2=0.85, mode=global)

θ1=0.9, θ2=0.85, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.980)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: cache (0.979)

[session\_000] Q5: cache (0.959)

[session\_001] Q1: cache (0.967)

[session\_001] Q2: cache (0.954)

[session\_002] Q1: cache (0.944)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: cache (0.981)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.982)

[session\_003] Q3: cache (0.962)

[session\_003] Q4: cache (0.976)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.977)

[session\_004] Q2: cache (0.958)

[session\_005] Q1: cache (0.920)

[session\_005] Q2: cache (0.910)

[session\_005] Q3: cache (0.914)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: cache (0.958)

[session\_006] Q1: llm (0.000)

[session\_006] Q2: llm (0.000)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: cache (0.979)

[session\_007] Q3: cache (0.975)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: cache (0.947)

[session\_008] Q1: cache (0.920)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: cache (0.913)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.987)

[session\_009] Q3: cache (0.980)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: cache (0.960)

[session\_010] Q3: cache (0.969)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.983)

[session\_011] Q3: cache (0.975)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: cache (0.966)

[session\_012] Q3: cache (0.950)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: cache (0.931)

[session\_013] Q1: cache (0.989)

[session\_013] Q2: cache (0.980)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.977)

[session\_014] Q3: cache (0.984)

[session\_014] Q4: cache (0.976)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: cache (0.948)

[session\_015] Q2: cache (0.959)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.974)

[session\_016] Q3: cache (0.957)

[session\_016] Q4: cache (0.955)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: cache (0.950)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 256d - Relaxed (θ1=0.85, θ2=0.8, mode=global)

θ1=0.85, θ2=0.8, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.980)

[session\_000] Q3: cache (0.946)

[session\_000] Q4: cache (0.939)

[session\_000] Q5: cache (0.959)

[session\_001] Q1: cache (0.967)

[session\_001] Q2: cache (0.954)

[session\_002] Q1: cache (0.944)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: cache (0.981)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.982)

[session\_003] Q3: cache (0.962)

[session\_003] Q4: cache (0.976)

[session\_003] Q5: cache (0.951)

[session\_004] Q1: cache (0.977)

[session\_004] Q2: cache (0.948)

[session\_005] Q1: cache (0.920)

[session\_005] Q2: cache (0.910)

[session\_005] Q3: cache (0.914)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: cache (0.973)

[session\_006] Q1: cache (0.952)

[session\_006] Q2: cache (0.909)

[session\_007] Q1: cache (0.885)

[session\_007] Q2: cache (0.878)

[session\_007] Q3: cache (0.889)

[session\_007] Q4: cache (0.911)

[session\_007] Q5: cache (0.906)

[session\_008] Q1: cache (0.920)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: cache (0.972)

[session\_008] Q4: cache (0.959)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.987)

[session\_009] Q3: cache (0.980)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: cache (0.974)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: cache (0.960)

[session\_010] Q3: cache (0.969)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: cache (0.971)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.983)

[session\_011] Q3: cache (0.975)

[session\_011] Q4: cache (0.948)

[session\_011] Q5: cache (0.937)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: cache (0.966)

[session\_012] Q3: cache (0.950)

[session\_012] Q4: cache (0.927)

[session\_012] Q5: cache (0.931)

[session\_013] Q1: cache (0.989)

[session\_013] Q2: cache (0.932)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.977)

[session\_014] Q3: cache (0.984)

[session\_014] Q4: cache (0.976)

[session\_014] Q5: cache (0.963)

[session\_015] Q1: cache (0.948)

[session\_015] Q2: cache (0.942)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.974)

[session\_016] Q3: cache (0.957)

[session\_016] Q4: cache (0.955)

[session\_016] Q5: cache (0.946)

[session\_017] Q1: cache (0.950)

[session\_017] Q2: cache (0.927)

[session\_017] Q3: llm (0.000)

==================================================

Testing 512d - Strict (θ1=0.95, θ2=0.9, mode=global)

θ1=0.95, θ2=0.9, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: llm (0.000)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: llm (0.000)

[session\_000] Q5: llm (0.000)

[session\_001] Q1: cache (0.968)

[session\_001] Q2: llm (0.000)

[session\_002] Q1: llm (0.000)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: llm (0.000)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.978)

[session\_003] Q3: llm (0.000)

[session\_003] Q4: llm (0.000)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.971)

[session\_004] Q2: llm (0.000)

[session\_005] Q1: llm (0.000)

[session\_005] Q2: cache (0.990)

[session\_005] Q3: llm (0.000)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: llm (0.000)

[session\_006] Q2: llm (0.000)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: llm (0.000)

[session\_007] Q3: llm (0.000)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: llm (0.000)

[session\_008] Q1: llm (0.000)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: llm (0.000)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: llm (0.000)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: llm (0.000)

[session\_010] Q3: llm (0.000)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.980)

[session\_011] Q3: llm (0.000)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: llm (0.000)

[session\_012] Q3: llm (0.000)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: llm (0.000)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.969)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: llm (0.000)

[session\_014] Q3: cache (0.978)

[session\_014] Q4: llm (0.000)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: llm (0.000)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: llm (0.000)

[session\_016] Q3: llm (0.000)

[session\_016] Q4: llm (0.000)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: llm (0.000)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 512d - Balanced (θ1=0.9, θ2=0.85, mode=global)

θ1=0.9, θ2=0.85, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.975)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: llm (0.000)

[session\_000] Q5: cache (0.942)

[session\_001] Q1: cache (0.961)

[session\_001] Q2: llm (0.000)

[session\_002] Q1: cache (0.920)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: llm (0.000)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.978)

[session\_003] Q3: llm (0.000)

[session\_003] Q4: cache (0.969)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.971)

[session\_004] Q2: llm (0.000)

[session\_005] Q1: llm (0.000)

[session\_005] Q2: cache (0.990)

[session\_005] Q3: cache (0.977)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: cache (0.942)

[session\_006] Q2: cache (0.944)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: cache (0.973)

[session\_007] Q3: cache (0.968)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: llm (0.000)

[session\_008] Q1: llm (0.000)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: cache (0.944)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: cache (0.975)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: llm (0.000)

[session\_010] Q3: cache (0.959)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.980)

[session\_011] Q3: cache (0.970)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: llm (0.000)

[session\_012] Q3: llm (0.000)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: llm (0.000)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.969)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.968)

[session\_014] Q3: cache (0.978)

[session\_014] Q4: cache (0.967)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: cache (0.931)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.970)

[session\_016] Q3: cache (0.950)

[session\_016] Q4: cache (0.947)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: cache (0.934)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 512d - Relaxed (θ1=0.85, θ2=0.8, mode=global)

θ1=0.85, θ2=0.8, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.975)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: cache (0.974)

[session\_000] Q5: cache (0.942)

[session\_001] Q1: cache (0.961)

[session\_001] Q2: cache (0.937)

[session\_002] Q1: cache (0.920)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: cache (0.974)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.978)

[session\_003] Q3: cache (0.948)

[session\_003] Q4: cache (0.969)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.971)

[session\_004] Q2: cache (0.940)

[session\_005] Q1: cache (0.898)

[session\_005] Q2: cache (0.888)

[session\_005] Q3: cache (0.890)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: cache (0.944)

[session\_006] Q1: cache (0.854)

[session\_006] Q2: llm (0.000)

[session\_007] Q1: cache (0.851)

[session\_007] Q2: llm (0.000)

[session\_007] Q3: cache (0.979)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: cache (0.956)

[session\_008] Q1: cache (0.899)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: cache (0.964)

[session\_008] Q4: cache (0.875)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: cache (0.975)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: cache (0.944)

[session\_010] Q3: cache (0.959)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.980)

[session\_011] Q3: cache (0.970)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: cache (0.919)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: cache (0.955)

[session\_012] Q3: cache (0.932)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: cache (0.909)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.969)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.968)

[session\_014] Q3: cache (0.978)

[session\_014] Q4: cache (0.967)

[session\_014] Q5: cache (0.950)

[session\_015] Q1: cache (0.931)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.970)

[session\_016] Q3: cache (0.950)

[session\_016] Q4: cache (0.947)

[session\_016] Q5: cache (0.934)

[session\_017] Q1: cache (0.934)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 768d - Strict (θ1=0.95, θ2=0.9, mode=global)

θ1=0.95, θ2=0.9, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: llm (0.000)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: llm (0.000)

[session\_000] Q5: llm (0.000)

[session\_001] Q1: cache (0.972)

[session\_001] Q2: llm (0.000)

[session\_002] Q1: llm (0.000)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: llm (0.000)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.979)

[session\_003] Q3: llm (0.000)

[session\_003] Q4: cache (0.971)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.972)

[session\_004] Q2: llm (0.000)

[session\_005] Q1: llm (0.000)

[session\_005] Q2: cache (0.990)

[session\_005] Q3: cache (0.979)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: llm (0.000)

[session\_006] Q2: llm (0.000)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: llm (0.000)

[session\_007] Q3: llm (0.000)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: llm (0.000)

[session\_008] Q1: llm (0.000)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: llm (0.000)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: llm (0.000)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: llm (0.000)

[session\_010] Q3: llm (0.000)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.982)

[session\_011] Q3: llm (0.000)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: llm (0.000)

[session\_012] Q3: llm (0.000)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: llm (0.000)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.971)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: llm (0.000)

[session\_014] Q3: cache (0.980)

[session\_014] Q4: llm (0.000)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: llm (0.000)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: llm (0.000)

[session\_016] Q3: llm (0.000)

[session\_016] Q4: llm (0.000)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: llm (0.000)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 768d - Balanced (θ1=0.9, θ2=0.85, mode=global)

θ1=0.9, θ2=0.85, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.977)

[session\_000] Q3: llm (0.000)

[session\_000] Q4: llm (0.000)

[session\_000] Q5: cache (0.949)

[session\_001] Q1: cache (0.965)

[session\_001] Q2: llm (0.000)

[session\_002] Q1: cache (0.929)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: llm (0.000)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.979)

[session\_003] Q3: llm (0.000)

[session\_003] Q4: cache (0.971)

[session\_003] Q5: llm (0.000)

[session\_004] Q1: cache (0.972)

[session\_004] Q2: llm (0.000)

[session\_005] Q1: cache (0.906)

[session\_005] Q2: llm (0.000)

[session\_005] Q3: cache (0.986)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: cache (0.965)

[session\_006] Q2: cache (0.950)

[session\_007] Q1: llm (0.000)

[session\_007] Q2: cache (0.976)

[session\_007] Q3: cache (0.972)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: cache (0.941)

[session\_008] Q1: cache (0.909)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: llm (0.000)

[session\_008] Q4: llm (0.000)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: cache (0.975)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: llm (0.000)

[session\_010] Q3: cache (0.959)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.982)

[session\_011] Q3: cache (0.974)

[session\_011] Q4: llm (0.000)

[session\_011] Q5: llm (0.000)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: cache (0.958)

[session\_012] Q3: llm (0.000)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: llm (0.000)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.973)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.971)

[session\_014] Q3: cache (0.980)

[session\_014] Q4: cache (0.971)

[session\_014] Q5: llm (0.000)

[session\_015] Q1: cache (0.933)

[session\_015] Q2: llm (0.000)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.973)

[session\_016] Q3: cache (0.956)

[session\_016] Q4: cache (0.954)

[session\_016] Q5: llm (0.000)

[session\_017] Q1: cache (0.941)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

Testing 768d - Relaxed (θ1=0.85, θ2=0.8, mode=global)

θ1=0.85, θ2=0.8, mode=global

[session\_000] Q1: llm (0.000)

[session\_000] Q2: cache (0.977)

[session\_000] Q3: cache (0.940)

[session\_000] Q4: cache (0.930)

[session\_000] Q5: cache (0.949)

[session\_001] Q1: cache (0.965)

[session\_001] Q2: cache (0.942)

[session\_002] Q1: cache (0.929)

[session\_002] Q2: llm (0.000)

[session\_002] Q3: cache (0.973)

[session\_003] Q1: llm (0.000)

[session\_003] Q2: cache (0.979)

[session\_003] Q3: cache (0.953)

[session\_003] Q4: cache (0.971)

[session\_003] Q5: cache (0.939)

[session\_004] Q1: cache (0.972)

[session\_004] Q2: cache (0.936)

[session\_005] Q1: cache (0.906)

[session\_005] Q2: cache (0.897)

[session\_005] Q3: cache (0.899)

[session\_005] Q4: llm (0.000)

[session\_005] Q5: llm (0.000)

[session\_006] Q1: cache (0.866)

[session\_006] Q2: cache (0.950)

[session\_007] Q1: cache (0.862)

[session\_007] Q2: llm (0.000)

[session\_007] Q3: cache (0.982)

[session\_007] Q4: llm (0.000)

[session\_007] Q5: cache (0.959)

[session\_008] Q1: cache (0.909)

[session\_008] Q2: llm (0.000)

[session\_008] Q3: cache (0.968)

[session\_008] Q4: cache (0.883)

[session\_008] Q5: llm (0.000)

[session\_009] Q1: llm (0.000)

[session\_009] Q2: cache (0.985)

[session\_009] Q3: cache (0.975)

[session\_009] Q4: llm (0.000)

[session\_009] Q5: llm (0.000)

[session\_010] Q1: llm (0.000)

[session\_010] Q2: cache (0.947)

[session\_010] Q3: cache (0.959)

[session\_010] Q4: llm (0.000)

[session\_010] Q5: llm (0.000)

[session\_011] Q1: llm (0.000)

[session\_011] Q2: cache (0.982)

[session\_011] Q3: cache (0.974)

[session\_011] Q4: cache (0.939)

[session\_011] Q5: cache (0.929)

[session\_012] Q1: llm (0.000)

[session\_012] Q2: cache (0.958)

[session\_012] Q3: cache (0.935)

[session\_012] Q4: llm (0.000)

[session\_012] Q5: cache (0.972)

[session\_013] Q1: cache (0.986)

[session\_013] Q2: cache (0.971)

[session\_014] Q1: llm (0.000)

[session\_014] Q2: cache (0.971)

[session\_014] Q3: cache (0.980)

[session\_014] Q4: cache (0.971)

[session\_014] Q5: cache (0.953)

[session\_015] Q1: cache (0.933)

[session\_015] Q2: cache (0.926)

[session\_016] Q1: llm (0.000)

[session\_016] Q2: cache (0.973)

[session\_016] Q3: cache (0.956)

[session\_016] Q4: cache (0.954)

[session\_016] Q5: cache (0.941)

[session\_017] Q1: cache (0.941)

[session\_017] Q2: llm (0.000)

[session\_017] Q3: llm (0.000)

==================================================

THRESHOLD COMPARISON

==================================================

256d - Strict: 25.4% hit rate, 19.2x speedup

256d - Balanced: 57.7% hit rate, 22.8x speedup

256d - Relaxed: 78.9% hit rate, 23.4x speedup

512d - Strict: 12.7% hit rate, 24.5x speedup

512d - Balanced: 40.8% hit rate, 18.0x speedup

512d - Relaxed: 63.4% hit rate, 19.6x speedup

768d - Strict: 15.5% hit rate, 20.6x speedup

768d - Balanced: 43.7% hit rate, 20.4x speedup

768d - Relaxed: 69.0% hit rate, 19.8x speedup

Markdown table:

| Dim | Setting | Hit rate | Speedup | LLM calls | Cache hits | Total |

|-----|---------|----------|---------|-----------|------------|-------|

| 256 | Strict | 25.4% | 19.2x | 53 | 18 | 71 |

| 256 | Balanced | 57.7% | 22.8x | 30 | 41 | 71 |

| 256 | Relaxed | 78.9% | 23.4x | 15 | 56 | 71 |

| 512 | Strict | 12.7% | 24.5x | 62 | 9 | 71 |

| 512 | Balanced | 40.8% | 18.0x | 42 | 29 | 71 |

| 512 | Relaxed | 63.4% | 19.6x | 26 | 45 | 71 |

| 768 | Strict | 15.5% | 20.6x | 60 | 11 | 71 |

| 768 | Balanced | 43.7% | 20.4x | 40 | 31 | 71 |

| 768 | Relaxed | 69.0% | 19.8x | 22 | 49 | 71 |

Saved: threshold\_dim\_comparison.json

Process finished with exit code 0